2016-02-08

Variational Theory for Optimization under Stochastic Ambiguity

Royset, Johannes O.

Variational Theory for Optimization under Stochastic Ambiguity

Johannes O. Royset

Operations Research Department
Naval Postgraduate School
joroyset@nps.edu

Roger J-B Wets

Department of Mathematics
University of California, Davis
rjbwets@ucdavis.edu

Abstract. Stochastic ambiguity provides a rich class of uncertainty models that includes those in stochastic, robust, risk-based, and semi-infinite optimization, and that accounts for both uncertainty about parameter values as well as incompleteness of the description of uncertainty. We provide a novel, unifying perspective on optimization under stochastic ambiguity that rests on two pillars. First, the paper models ambiguity by decision-dependent collections of cumulative distribution functions viewed as subsets of a metric space of upper semicontinuous functions. We derive a series of results for this setting including estimates of the metric, the hypo-distance, and a new proof of the equivalence with weak convergence. Second, we utilize the theory of lopsided convergence to establish existence, convergence, and approximation of solutions of optimization problems with stochastic ambiguity. For the first time, we estimate the lop-distance between bifunctions and show that this leads to bounds on the solution quality for problems with stochastic ambiguity. Among other consequences, these results facilitate the study of the "price of robustness" and related quantities.

Keywords: stochastic ambiguity, robust optimization, lopsided convergence, lop-distance rate of convergence, price of robustness, weak convergence

AMS Classification: 90C15, 90C34, 90C47, 65K10

Date: February 8, 2016

1 Introduction

Optimization under stochastic ambiguity accounts for uncertainty in parameters as well as the fact that *models of uncertainty* might also be imprecise. The goal is to minimize by choice of decision variables the worst-case value of an objective function over a set of probability measures. A measure models the uncertainty about parameter values and the set captures the ambiguity about the correct measure. Optimization under stochastic ambiguity includes as special cases robust [5, 7, 12], stochastic [9, 18, 40], semi-infinite [22, 14, 41], and risk-based [21, 38, 28, 40] optimization. The references provide a glimpse into a vast literature on these subjects, where applications in finance are especially prevalent; see for example [10, 20].

In this paper, we provide a novel formulation in terms of cumulative distribution functions for this broad class of optimization problems and examine for the first time the application of the variational theory of lopsided convergence to establish existence, convergence, and approximation of solutions for optimization problems with stochastic ambiguity. Specifically, the paper provides a pathway to proving convergence and rate of convergence for approximation-based algorithms as well as constructing estimates of the price of ambiguity and robustness, i.e., the increase in cost faced by a decision maker uncertain about parameter values. In the process of developing result in this context, we obtain, with little overhead, results of significance more broadly for min-sup problems and probability theory.

There are three interconnected components to our approach. First, we formulate optimization under stochastic ambiguity in terms of *cumulative distribution functions* on $I\!R^m$, instead of probability measures, which are then treated as a subset of the metric space of upper semicontinuous functions equipped with the hypo-distance. This results in a novel treatment of distribution functions on $I\!R^m$, including the definition of convexity on this metric space and convenient estimates of the hypo-distance. Moreover, we provide an explicit statement of the fact that convergence of distribution functions in the hypo-distance is equivalent to weak convergence, with a new proof exclusive relying on elementary concepts from variational analysis; the result is implicit in [37, 36]. Of course, there are numerous metrics available for spaces of probability measures including the Levy-Prokhorov metric, which also characterizes weak convergence in the present context, the Wasserstein metrics, which require finite moments and is stronger than weak convergence, and the total variation metric, which is also stronger than weak convergence. This paper is the first attempt to use the hypo-distance as a metric for distribution functions in the context of optimization under stochastic ambiguity. As we see below, it is a promising alternative because of the natural interpretation of the hypo-distance, its equivalence with weak convergence, and the fact that it can be estimated in a reasonably convenient manner. Moreover, it can easily handle any finite dimension, i.e., any number of uncertain parameters. In fact, its extension to infinite-dimensional spaces appears clear (see the ideas in [36]), but such possibilities are beyond the scope of the present paper.

Second, we adopt the variational theory of *lopsided convergence* of bifunctions (i.e., functions taking two inputs) to examine convergence of approximations of optimization problems with stochastic ambiguity. The notion originated with [2] and later was modified and extended in [15, 16, 34, 33]. We show that lopsided convergence can be used to establish the existence of solutions of optimization problems with stochastic ambiguity as well as to prove the convergence of solutions of approximate problems to those of an original problem under mild assumptions.

Third, we estimate, for the first time, the *lop-distance* between bifunctions, a quantification of lopsided convergence developed in [33]. Utilizing these estimates, the paper bounds the difference between optimal solutions and optimal values of two optimization problems with stochastic ambiguity. The results provide a new quantitative theory of approximation for such problems.

Throughout, we aim to keep assumptions at the minimum. In particular, we do not insist universally that the feasible set of decision variables and the set of cumulative probability functions under consideration are compact. Neither do we require convexity and/or concavity. We refer to the extensive literature on robust, stochastic, and risk-based optimization for more specialized results; see the first paragraph for some references.

The paper proceeds in Section 2 by giving problem statements and illustrative examples. Section 3 develops the foundations for studying distribution functions in the present context. Section 4 reviews the notion of lopsided convergence, states the main consequences, and provides specific results for optimization under stochastic ambiguity. Section 5 offers estimates of the lop-distance as well as solution quality for min-sup problems generally. We also give concrete examples in the context of optimization under stochastic ambiguity.

2 Problem Formulation and Examples

We start by defining the problem of optimization under stochastic ambiguity, which is then followed by several examples illustrating the breadth of applications. The section ends with a reformulation in terms of cumulative distribution functions. Although extensions are possible and in fact trivial in some cases, we limit the scope to optimization of an n-dimensional vector of decision variables in the presence of uncertainty about m parameters. This enables us to avoid much of the topological and other technical considerations needed in the more general cases.

We let $X \subset I\!R^n$ be a nonempty set of feasible decision vectors, possibly being the whole of $I\!R^n$. The set of all probability measures on $(I\!R^m, \mathcal{B}_{I\!R^m})$ is denoted by $\mathcal{M}$, where $\mathcal{B}_{I\!R^m}$ is the Borel σ-algebra on $I\!R^m$. Stochastic ambiguity models uncertainty about the choice of probability measure by considering a "worst-case" probability measure over a family of measures. We allow this family to depend on the decision vector $x \in X$ and therefore define the set-valued mapping

$$\mathcal{P} : X \rightrightarrows \mathcal{M}, \text{ with } \mathcal{P}(x) \neq \emptyset \text{ for all } x \in X.$$

The set of decision vectors and probability measures that needs to be considered is therefore $\{(x, P) \in X \times \mathcal{M} \;:\; P \in \mathcal{P}(x)\}$. On this set, an objective function φ is finite. This leads to the following optimization problem with stochastic ambiguity:

$$\min_{x \in X} \sup_{P \in \mathcal{P}(x)} \varphi(x, P). \tag{1}$$

The problem arises broadly as exemplified next.

2.1 Illustrations

The optimization problem in (1) provides practitioners the ability to model decision making under uncertainty accounting for possibly incomplete, uncertain, and decision-dependent descriptions of unknown parameters. Three examples illustrate some possibilities.

Example 1: Expectation Minimization. For some $\psi : X \times I\!R^m \to I\!R$, with $\psi(x, \cdot)$ integrable with respect to P for all $x \in X$ and $P \in \mathcal{P}(x)$, the problem[1]

$$\min_{x \in X} \sup_{P \in \mathcal{P}(x)} E_P[\psi(x, \boldsymbol{\xi})] := \int \psi(x, \xi) dP(\xi)$$

is a special case of (1), with $\varphi(x, P) = E_P[\psi(x, \boldsymbol{\xi})]$. It captures stochastic optimization under distributional uncertainty (most often considered with $\mathcal{P}$ being independent of x), stochastic programs with

[1]Random vectors are indicated by boldface font and their realizations by regular font.

decision-dependent probability measure (for example with $\mathcal{P}(x)$ being a singleton for every $x \in X$), and certain Stackelberg games; see for example [17, 19, 25, 39].

Example 2: Risk Minimization. Suppose that $P_0 \in \mathcal{M}$ is a reference probability measure and $\mathcal{L}^2(\mathbb{R}^m) := \{g : \mathbb{R}^m \to \mathbb{R} \ : \ \int g(\xi)^2 dP_0(\xi) < \infty\}$. Let $\mathcal{R}$ be a measure of risk on $\mathcal{L}^2(\mathbb{R}^m)$ for a (risk-averse) decision maker that aims to safeguard against undesirable outcomes. If $\psi(x, \cdot) \in \mathcal{L}^2(\mathbb{R}^m)$ for all $x \in X$, then the goal might be to minimize $\mathcal{R}(\psi(\cdot, \boldsymbol{\xi}))$ over X. If $\mathcal{R}$ is regular, monotone, and positively homogeneous (which is equivalent to coherency [1]), then $\mathcal{R}(\psi(x, \boldsymbol{\xi})) = \sup_{P \in \mathcal{P}} \int \psi(x, \xi) dP(\xi)$, where $\mathcal{P}$ is the set of probability measures on $\mathbb{R}^m$ absolutely continuous with respect to P_0 that each has a density in the effective domain of the conjugate of $\mathcal{R}$; see for example [35, 13, 28, 25]. Consequently, we have returned to a problem of the form in Example 1, but with $\mathcal{P}$ independent of x. Particular instances of such risk measures are first- and second-order superquantile risk measures[2] [27, 26].

Example 3: Robust Optimization. We again start from Example 1, but now specialize in a different direction. For some set-valued mapping $\Xi : X \rightrightarrows \mathbb{R}^m$, with $\Xi(x) \neq \emptyset$ for all $x \in X$, let

$$\mathcal{P}(x) = \{P \in \mathcal{M} \ : \ P(\xi) = 1 \text{ for some } \xi \in \Xi(x)\} \text{ for } x \in X.$$

Then, expectation minimization of Example 1 simplifies to $\min_{x \in X} \sup_{\xi \in \Xi(x)} \psi(x, \xi)$, which encapsulates many robust optimization and (generalized) semi-infinite programming problems; see for example [22, 41, 5, 7, 12].

2.2 Formulation using Distribution Functions

Although one could adopt a metric on the space of probability measures and then proceed with analysis of (1), we here explore for the first time another possibility centered on distribution functions. Practitioners usually think about uncertainty in terms of a random variable $\boldsymbol{\xi}$ with some distribution function that is more or less known. Consequently, it is natural to formulate optimization under stochastic ambiguity in terms of distribution functions that must be selected from a set of "plausible" functions. Moreover, a natural and geometrically intuitive metric, the hypo-distance to be introduced in Subsection 3.1, is available to quantify the distance between two distribution functions on $\mathbb{R}^m$. The convergence induced by this metric is equivalent to weak convergence of distribution functions as seen in the next section. Thus, the topology generated by the metric seems to be an interesting possibility.

We start by recalling some well-known facts about distribution functions on $\mathbb{R}^m$. Every probability measure P on $(\mathbb{R}^m, \mathcal{B}_{\mathbb{R}^m})$ defines a *distribution function* $F : \mathbb{R}^m \to [0, 1]$ through

$$F(\xi) = P(S_\xi) \text{ for } \xi \in \mathbb{R}^m, \text{ where } S_\xi := \{\zeta \in \mathbb{R}^m \ : \ \zeta \leq \xi\}.$$

A distribution function F is nondecreasing, i.e., $F(\zeta) \leq F(\xi)$ for $\zeta \leq \xi$, it is continuous from above[3], it satisfies $F(\xi^\nu) \to 0$ whenever one of the components of ξ^ν tends to $-\infty$, with the others held fixed,

[2] A superquantile of a random variable is called conditional value-at-risk and average value-at-risk in the finance literature. We here adopt the application-neutral terminology proposed in [23].

[3] $F(\xi^\nu) \to F(\xi)$ if $\xi^\nu = \xi + \zeta^\nu$ and $\zeta_i^\nu \downarrow 0$ for all $i = 1, 2, ..., m$.

and $F(\xi^\nu) \to 1$ if $\xi_i^\nu \to \infty$ for all i, and has $\Delta_A F \geq 0$ for every rectangle[4] A, where

$$\Delta_A F := \sum_{j=1}^{2^m} \left(\operatorname{sgn}_A v^j\right) F(v^j),$$

with v^j, $j = 1, ..., 2^m$ being the vertices of A and $\operatorname{sgn}_A v^j = 1$ if the number of components v_i^j at a lower bound of A is even and $\operatorname{sgn}_A v^j = -1$ if the number is odd. For example, for $m = 2$, we have that $\Delta_A F = F(v^1) - F(v^2) - F(v^3) + F(v^4)$, where v^1 is the upper right vertex, v^2 and v^3 are the upper left and lower right vertices, and v^4 is the lower left vertex.

For every $F : I\!R^m \to I\!R$ with these properties (nondecreasing, continuity from above, limits, and $\Delta_A F \geq 0$), there exists a unique probability measure P on $(I\!R^m, \mathcal{B}_{I\!R^m})$ such that $P(A) = \Delta_A F$ for rectangles A and $P(S_\xi) = F(\xi)$ for all $\xi \in I\!R^m$. Consequently, nothing is lost by proceeding with a reformulation of (1) in terms of distribution functions.

For the remainder of the paper, we therefore consider the *stochastic ambiguity problem*

$$(\text{SAP}): \quad \min_{x\in X} \sup_{F\in\mathcal{F}(x)} \varphi(x, F), \text{ where } \mathcal{F} : X \rightrightarrows \mathcal{D} \text{ has } \mathcal{F}(x) \neq \emptyset \text{ for all } x \in X,$$

where X is a nonempty subset of $I\!R^n$, the set of distribution functions

$$\mathcal{D} := \{F : I\!R^m \to [0,1] \ : \ \text{for some } P \in \mathcal{M}, F(\xi) = P(S_\xi) \ \forall \xi \in I\!R^m\},$$

and the objective function

$$\varphi : Z \to I\!R, \text{ with } Z := \Big\{(x, F) \in X \times \mathcal{D} \ : \ F \in \mathcal{F}(x)\Big\}.$$

For $x \in X$, $\mathcal{F}(x)$ is an *ambiguity set* specifying the collection of distribution functions that needs to be considered under decision x. The required nonemptyness of $\mathcal{F}(x)$ for all $x \in X$ implies that we rule out the pathological case where there exists an $x \in X$ for which the "payoff" is $-\infty$ even in the worst case.

An x that achieves the minimum in (SAP) is called a *minsup-point* of (SAP) and is denoted by $\operatorname{argmin}_X \sup_{\mathcal{F}} \varphi$. The optimal value of (SAP) is denoted by $\inf_X \sup_{\mathcal{F}} \varphi$ and is called its *minsup-value.* The function φ might be defined and finite-valued outside Z, but that will be immaterial to the following treatment. Since φ depends critically on X and $\mathcal{F}$, we make this dependence explicit and often denote it by $\varphi_{X\mathcal{F}}$ and refer to it as a bifunction due to its dependence on two inputs x and F. The minsup-points and minsup-value of (SAP) are also called minsup-points and minsup-value of $\varphi_{X\mathcal{F}}$, respectively.

3 Foundations for Distribution Functions

The formulation of optimization under stochastic ambiguity in terms of distribution functions requires us to develop the necessary mathematical tools for analyzing the inner maximization over such functions. We view distribution functions as a subset of a space of upper semicontinuous functions as described next. Following introductory definitions, we develop a series of results that facilitate understanding and analysis of distribution functions in this context. Since viewing distribution functions as a subset of the upper semicontinuous functions might be beneficial in other contexts too, we believe that this section contributes beyond the present context.

[4] A rectangle in $I\!R^m$ is of the form $A = \{\xi \ : \ a_i < \xi_i \leq b_i, i = 1, 2, ..., m\}$ for real a_i and b_i.

3.1 Space of Upper Semicontinuous Functions

To facilitate analysis of distribution functions, we view them as a subset of a larger class of functions. We recall that a function $g : I\!R^m \to \overline{I\!R} := I\!R \cup \{-\infty, \infty\}$ is extended real-valued upper semicontinuous (usc) if for every $\xi \in I\!R^m$ and $\xi^\nu \to \xi$, $\limsup g(\xi^\nu) \leq g(\xi)$. Let the set of all such functions, excluding the function that is identical to $-\infty$, be denoted by usc-fcns($I\!R^m$). If $F \in \mathcal{D}$, $\xi^\nu \to \xi$, and $\zeta_i^\nu = \max\{\xi_i, \xi_i^\nu\}$ for all i, then $F(\xi^\nu) \leq F(\zeta^\nu)$ by the nondecreasing property of F. Since $F(\zeta^\nu) \to F(\xi)$ by the continuity-from-above property, it is clear that F is usc and, thus $\mathcal{D} \subset$ usc-fcns($I\!R^m$).

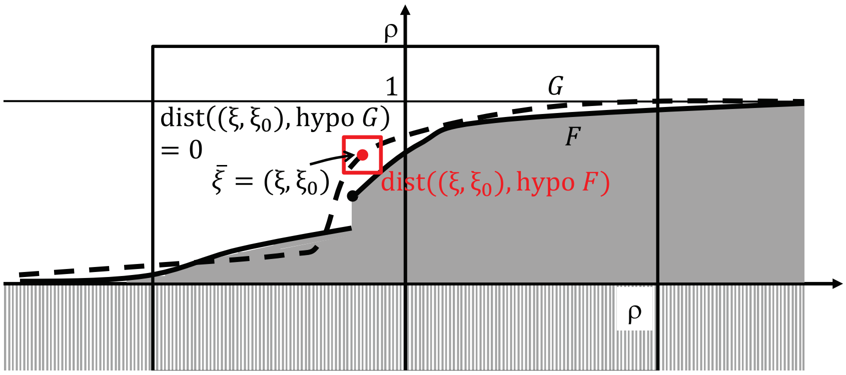

Figure 1: Hypo-distance between two distribution functions F (solid line) and G (dashed line).

We embed the space usc-fcns($I\!R^m$) with the *hypo-distance* $d\!l^{\mathrm{h}}$, which quantifies the distance between usc functions in terms of a distance between their hypo-graphs. As usual, for $g \in$ usc-fcns($I\!R^m$), the hypo-graph of g is

$$\operatorname{hypo} g := \big\{(\xi, \xi_0) \in I\!R^m \times I\!R \;:\; g(\xi) \geq \xi_0\big\}.$$

Clearly, hypo g is a subset of $I\!R^{m+1}$ and a possibility would be to adopt the Euclidean norm on $I\!R^{m+1}$ when measuring distances between points in $I\!R^{m+1}$, which is indeed the approach followed in [29, Chapters 4 and 7]. Here we choose a different setup and make a distinction between the last component of vectors in $I\!R^{m+1}$ and the other components. Thus, for $I\!R^m \times I\!R$ we adopt the norm

$$\|(\xi, \xi_0)\|_{\mathbb{S}} := \max\{\|\xi\|, |\xi_0|\} \text{ for } (\xi, \xi_0) \in I\!R^m \times I\!R. \tag{2}$$

There are two main motivations for this approach. First, bounds on the hypo-distance between two functions simplify under this choice and, second, it facilitates generalizations to situations when $I\!R^m$ is replaced by a general metric space. We benefit from the former in this paper, but postpone realizing the benefit from the latter to a subsequent paper. A ball under $\|\cdot\|_{\mathbb{S}}$, with radius r centered at $\bar\xi = (\xi, \xi_0) \in I\!R^m \times I\!R$, is denoted by

$$\mathbb{S}(\bar\xi, r) := \big\{\bar\zeta \in I\!R^m \times I\!R \;:\; \|\bar\xi - \bar\zeta\|_{\mathbb{S}} \leq r\big\}$$

and is actually a "hyper-cylinder," as hinted to by the symbol $\mathbb{S}$. Figure 1 show $\mathbb{S}(0, \rho)$ for the case $m = 1$. The distance between $\bar\xi \in I\!R^m \times I\!R$ and a set $S \subset I\!R^m \times I\!R$ is given by

$$\operatorname{dist}\big(\bar\xi, S\big) := \inf\Big\{\|\bar\xi - \bar\zeta\|_{\mathbb{S}} \;:\; \bar\zeta \in S\Big\}. \tag{3}$$

We are then in a position to define, for any $g, g' \in$ usc-fcns$(\mathbb{R}^m)$, the hypo-distance as

$$d\!l^{\mathrm{h}}(g, g') := \int_0^\infty d\!l^{\mathrm{h}}_\rho(g, g') e^{-\rho} d\rho,$$

where the *ρ-hypo-distance*

$$d\!l^{\mathrm{h}}_\rho(g, g') := \max\Big\{ \big| \operatorname{dist}\big(\bar{\xi}, \operatorname{hypo} g\big) - \operatorname{dist}\big(\bar{\xi}, \operatorname{hypo} g'\big) \big| \;:\; \|\bar{\xi}\|_{\mathbb{S}} \leq \rho \Big\}, \quad \text{for } \rho \geq 0.$$

Figure 1 illustrates the situation and especially the hypo-distance between distribution functions F and G. We observe that $d\!l^{\mathrm{h}}_\rho$ relates to the classical Pompeiu-Hausdorff distance, but uses "truncation" to handle the unboundedness of $\operatorname{hypo} g$ and $\operatorname{hypo} g'$. For further connections we reference [29, Exer. 7.60, Prop. 7.61]. At the end of the section, we give specific estimates of the hypo-distance between distribution functions beyond the immediate fact that $d\!l^{\mathrm{h}}(F, G) \leq 1$ for all $F, G \in \mathcal{D}$.

The hypo-distance is a metric on usc-fcns$(\mathbb{R}^m)$ and induces the *hypo-topology* (sometimes called the Attouch-Wets topology). In fact, we deduce from [29, Theorem 7.58] and [32, Corollary 3.6] that $(\text{usc-fcns}(\mathbb{R}^m), d\!l^{\mathrm{h}})$ is a complete separable metric (Polish) space. Every ball $\{g \in \text{usc-fcns}(\mathbb{R}^m) : d\!l^{\mathrm{h}}(f, g) \leq r\}$ in this space is compact as can be deduced from [29, Theorem 7.58]. (The difference in norm on $\mathbb{R}^m \times \mathbb{R}$ between the present paper and [29] is immaterial as they generate the same topology.)

We say that functions $g^\nu \in$ usc-fcns$(\mathbb{R}^m)$ *hypo-converge* to a function $g \in$ usc-fcns$(\mathbb{R}^m)$ if $d\!l^{\mathrm{h}}(g^\nu, g) \to 0$. Let $\mathbb{N} := \{1, 2, ..., \}$. By Theorem 7.58 in [29], g^ν hypo-converges to g if and only if $\operatorname{hypo} g^\nu$ set-converges[5] to $\operatorname{hypo} g$. A well-known convenient characterization of hypo-convergence is given next [29, Equation 7(9)].

3.1 Proposition (hypo-convergence for usc functions) *The functions $g^\nu \in$ usc-fcns$(\mathbb{R}^m)$ hypo-converge to $g \in$ usc-fcns$(\mathbb{R}^m)$ if and only if*

(i) for every $\xi^\nu \to \xi$, $\operatorname{limsup} g^\nu(\xi^\nu) \leq g(\xi)$;

(ii) for every ξ, there exists a sequence $\xi^\nu \to \xi$ such that $\operatorname{liminf} g^\nu(\xi^\nu) \geq g(\xi)$.

Due to its implication for the maximization-portion of (SAP), we also briefly discuss convexity of subsets of usc-fcns$(\mathbb{R}^m)$ and of functionals defined on such subsets. Although usc-fcns$(\mathbb{R}^m)$ is not a linear space, it is a pointed cone and the following definition, given here for the first time, is permissible under the usual conventions about extended real-valued algebra[6].

3.2 Definition (convexity on usc-fcns$(\mathbb{R}^m)$) *A set $S \subset$ usc-fcns$(\mathbb{R}^m)$ is convex if*

$$\text{for every } g, g' \in S \text{ and } \lambda \in [0,1],\; \lambda g + (1-\lambda) g' \in S.$$

A functional $\psi : S \to \mathbb{R}$ is convex if $S \subset$ usc-fcns$(\mathbb{R}^m)$ is a convex set and

$$\text{for every } g, g' \in S \text{ and } \lambda \in [0,1],\; \psi(\lambda g + (1-\lambda) g') \leq \lambda \psi(g) + (1-\lambda)\psi(g').$$

[5] We recall that the outer limit of a sequence of sets $\{A^\nu\}_{\nu \in \mathbb{N}}$, denoted by $\operatorname{limsup} A^\nu$, is the collection of points y to which a subsequence of $\{y^\nu\}_{\nu \in \mathbb{N}}$, with $y^\nu \in A^\nu$, converges. The inner limit, denoted by $\operatorname{liminf} A^\nu$, is the points to which a sequence of $\{y^\nu\}_{\nu \in \mathbb{N}}$, with $y^\nu \in A^\nu$, converges. If both limits exist and are identical to A, we say that $\{A^\nu\}_{\nu \in \mathbb{N}}$ set-converge to A. We retain this terminology for subsets of any metric space; see [4, 29].

[6] In the context of usc, we let $a + \infty = \infty$, $a + (-\infty) = -\infty$ for all $a \in \mathbb{R} \cup \{\infty\}$, $(-\infty) + \infty = (-\infty) + (-\infty) = -\infty$, $0 \cdot \infty = 0 \cdot (-\infty) = 0$, and $b \cdot \infty = (-b) \cdot (-\infty) = \infty$ and $(-b) \cdot \infty = b \cdot (-\infty) = -\infty$ for all $b > 0$.

Obviously, this definition has the usual implication that every local minimizer of a convex functional ψ over a convex set S is a global minimizer. We note that usc-fcns($I\!R^m$) is not a convex set as it contains functions g, g' with $(1/2)g + (1/2)g' \equiv -\infty$. However, for example, any set of the form $\{g \in \text{usc-fcns}(I\!R^m) \ : \ g(\xi) > -\infty \text{ for } \xi \in A \text{ and } g(x) = -\infty \text{ for } \xi \notin A\}$, with $A \subset I\!R^m$ closed and nonempty, is convex.

In view of Definition 3.2, the subsequent discussion, and the properties of distribution functions, $\mathcal{D}$ is a convex set. Moreover, the maximization problem in (SAP) is a convex problem for a given $x \in X$ if $\mathcal{F}(x)$ is a convex set and $-\varphi(x, \cdot)$ is convex on $\mathcal{F}(x)$. A concrete example follows next.

Example 4: Convexity under Moment Information. For distribution functions on $I\!R$, a particular choice of ambiguity set restricts the considerations to distributions with moments equal to $a_k(x)$, $k = 1, 2, ..., K$, i.e.,

$$\mathcal{F}(x) = \left\{ F \in \mathcal{D} \ : \ \int \xi^k dF(\xi) = a_k(x), \ k = 1, 2, ..., K \right\}.$$

Clearly, $\mathcal{F}(x)$ is a convex set due to the linearity of the integral.

3.2 Connections to Weak Convergence and Other Properties

We next examine properties of the metric space (usc-fcns($I\!R^m$), $d\!l^{\mathrm{h}}$), especially related to the subset $\mathcal{D}$ of distribution functions. In addition to be central for the subsequent development, these results are also of independent interest. We recall that probability measures $P^\nu \in \mathcal{M}$ converge *weakly* to a probability measure $P \in \mathcal{M}$ if and only if $\limsup P^\nu(A) \leq P(A)$ for all closed sets $A \subset I\!R^m$. This convergence takes place if and only if the corresponding distribution functions F^ν and F converge pointwise at all points of continuity of F. The distribution functions are then said to also converge weakly. The following result is implicit in [37, 36], where the development is more abstract dealing with probability semicontinuous measures on closed sets. Here, we provide for the first time an explicit statement for the present context and give a new simplified proof that only relies on Proposition 3.1.

3.3 Theorem (hypo-convergence equivalent to weak convergence) *For distribution functions $F^\nu, F \in \mathcal{D}$ we have that*

$$d\!l^{\mathrm{h}}(F^\nu, F) \to 0 \text{ if and only if } F^\nu \text{ converges weakly to } F.$$

Proof. Let the probability measures $P^\nu, P \in \mathcal{M}$ correspond to F^ν, F, i.e., $F^\nu(\xi) = P^\nu(S_\xi)$ and $F(\xi) = P(S_\xi)$ for all $\xi \in I\!R^m$. First, suppose that P^ν converges weakly to P. We utilize Proposition 3.1. For any $\xi \in I\!R^m$, S_ξ is closed and therefore $\limsup F^\nu(\xi) = \limsup P^\nu(S_\xi) \leq P(S_\xi) = F(\xi)$. Let $\xi^\nu \to \xi$ and $\varepsilon > 0$ be arbitrary. By continuity from above of F, there exists a $\zeta \in I\!R^m$, with $\zeta_i > 0$ for all $i = 1, ..., m$, such that $F(\xi + \zeta) \leq F(\xi) + \varepsilon$. Moreover, there exists a $\bar{\nu} \in I\!N$ such that $\xi^\nu \leq \xi + \zeta$ for all $\nu \geq \bar{\nu}$. Consequently, for all $\nu \geq \bar{\nu}$, we have by the monotonicity of F^ν that $F^\nu(\xi^\nu) \leq F^\nu(\xi + \zeta)$. This implies that

$$\limsup F^\nu(\xi^\nu) \leq \limsup F^\nu(\xi + \zeta) \leq F(\xi + \zeta) \leq F(x) + \varepsilon.$$

Since ε is arbitrary, part (i) of Proposition 3.1 holds. Next, let $\xi \in I\!R^m$ be arbitrary. Pick $\zeta \in I\!R^m$, with $\zeta_i > 0$ for all $i = 1, ..., m$. Then, there exists an open set A with $S_\xi \subset A \subset S_{\xi+\zeta}$. Since A^c (the complement of A) is closed, $\limsup P^\nu(A^c) \leq P(A^c)$ and therefore $\liminf P^\nu(A) \geq P(A)$. Monotonicity implies that

$$\liminf F^\nu(\xi+\zeta) = \liminf P^\nu(S_{\xi+\zeta}) \geq \liminf P^\nu(A) \geq P(A) \geq P(S_\xi) = F(\xi). \tag{4}$$

Now, let $\varepsilon > 0$, $\nu_0 = 0$, $\zeta^\mu \to 0$, with $\zeta_i^\mu > 0$ for all $i = 1, ..., m$ and $\mu \in I\!N$. For all μ, there exists by (4) a $\nu_\mu > \nu_{\mu-1}$ such that for all $\nu \geq \nu_\mu$, $F^\nu(\xi + \zeta^\mu) \geq F(\xi) - \varepsilon$. Construct the sequence $\{\xi^\nu\}_{\nu \in N}$ as follows. For $\nu_1 \leq \nu \leq \nu_2$, set $\xi^\nu = \xi + \zeta^1$. For $k = 2, 3, ...$, set $\xi^\nu = \xi + \zeta^k$ for $\nu_k < \nu \leq \nu_{k+1}$. By construction, $F^\nu(\xi^\nu) \geq F(\xi) - \varepsilon$. Moreover, $\xi^\nu \to \xi$ and part (ii) of Proposition 3.1 holds.

Second, we consider the converse and suppose that $d\!l^{\mathrm{h}}(F^\nu, F) \to 0$. We directly obtain from part (i) of Proposition 3.1 that $\limsup F^\nu(\xi) \leq F(\xi)$ for all $\xi \in I\!R^m$. Let $\xi \in I\!R^m$ be an arbitrary point at which F is continuous. We next establish that $\liminf F^\nu(\xi) \geq F(\xi)$. Let $\varepsilon > 0$. By continuity of F at ξ there exists a $\delta > 0$ such that $F(\zeta) \geq F(\xi) - \varepsilon$ for all $\zeta \in I\!R^m$ satisfying $\|\zeta - \xi\| \leq \delta$. Let ζ be one such point with $\zeta_i < \xi_i$ for all $i = 1, 2, ..., m$. Part (ii) of Proposition 3.1 implies that there exists a sequence $\zeta^\nu \to \zeta$ such that $\liminf F^\nu(\zeta^\nu) \geq F(\zeta)$. For this sequence, there exists a $\bar\nu \in I\!N$ such that $\zeta^\nu \leq \xi$ for all $\nu \geq \bar\nu$. The monotonicity of F^ν gives that $F^\nu(\xi) \geq F^\nu(\zeta^\nu)$ for all $\nu \geq \bar\nu$. Combining these results, we obtain that

$$\liminf F^\nu(\xi) \geq \liminf F^\nu(\zeta^\nu) \geq F(\zeta) \geq F(\xi) - \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, we conclude that $F^\nu(\xi) \to F(\xi)$ for ξ a point of continuity of F. □

Example 5: Hypo-Convergence of Empirical Distribution. If $\{\boldsymbol{\xi}^\nu\}_{\nu \in N}$ is a sequence of independent and identically distributed random vectors with values in $I\!R^m$ that is defined on a probability space $(\Omega, \mathcal{A}, \mu)$ and F is the distribution function of $\boldsymbol{\xi}^1$, then the empirical distribution functions[7]

$$F^\nu(\cdot, \omega) = \frac{1}{\nu}\sum_{j=1}^{\nu} I(\xi^j(\omega) \leq \cdot) \text{ hypo-converge to } F \text{ for } \mu\text{-almost every } \omega \in \Omega.$$

This fact is obvious in view of Theorem 3.3 since weak convergence of empirical measures holds in this case; see for example Theorem 11.4.1 in [11]. Of course, other conclusions are also available for empirical measures, but we here include this fact as an illustration of hypo-convergence.

In addition to empirical distribution functions, there are numerous other paths to constructing arbitrarily accurate approximations of a distribution function in the sense of $d\!l^{\mathrm{h}}$. In particular, we deduce from [32, Theorem 3.5] that epi-splines, a particular class of piecewise polynomial functions, can be employed for this purpose. A concrete example would be piecewise constant functions defined on a rectangular partition of $I\!R^m$.

It is well-known that the weak limit of a sequence of distribution functions might not be a distribution function. However, tightness ensures this property as recalled next.

3.4 Definition (tightness) *A subset $\mathcal{S} \subset \mathcal{D}$ is* tight *if*

for all $\varepsilon > 0$, there exists a rectangle A such that $\Delta_A F \geq 1 - \varepsilon$ for all $F \in \mathcal{S}$.

3.5 Proposition (convergence to distribution function) *If $\{F^\nu\}_{\nu \in N} \subset \mathcal{D}$ is tight, then the following hold:*

[7] We use the notation $I(\xi' \leq \xi) = 1$ if $\xi' \leq \xi$ and 0 otherwise.

(i) *There exists a subsequence* $\{F^{\nu_k}\}_{k\in\mathbb{N}}$ *and an* $F \in \mathcal{D}$ *such that* $d\!l^{\mathrm{h}}(F^{\nu_k}, F) \to 0$ *as* $k \to \infty$.

(ii) *If* $F : \mathbb{R}^m \to \mathbb{R}$ *is the limit of* $\{F^\nu\}_{\nu\in\mathbb{N}}$*, i.e.,* $d\!l^{\mathrm{h}}(F^\nu, F) \to 0$*, then* $F \in \mathcal{D}$.

Proof. In view of Theorem 3.3, this follows by standard results in probability theory; see for example [8, Theorem 29.3] and its corollary. □

Generally, tightness is closely related to compactness as is well known from Prokhorov's theorem. Next we state the connection in the present context, arguing directly from Proposition 3.5. First, however, we establish some notation. We denote by $\mathrm{cl}\, A$ the closure of a subset A of a topological space. In any metric space (M, d), we let the ball $\mathbb{B}(u, r) := \{u' \in M \;:\; d(u, u') \le r\}$. Moreover, $\mathbb{B} := \mathbb{B}(0, 1)$ and $r\mathbb{B} := \mathbb{B}(0, r)$ whenever the metric space contains a point 0. For $\mathbb{R}^k$, we use the usual Euclidean norm if not explicitly stated otherwise. We have already encountered an exception in the case of $\mathbb{R}^m \times \mathbb{R}$, which was given the norm $\|\cdot\|_{\mathbb{S}}$ with balls $\mathbb{S}(\bar\xi, r)$; see Subsection 3.1. Still, we let $r\mathbb{S} := \mathbb{S}(0, r)$.

3.6 Proposition (compactness and tightness) *For a set* $\mathcal{S} \subset \mathcal{D}$ *of distributions function, we have that*

(i) *if* $\mathcal{S}$ *is compact, then* $\mathcal{S}$ *is tight;*

(ii) *if* $\mathcal{S}$ *is tight, then* $\mathrm{cl}\,\mathcal{S}$ *is compact and contained in* $\mathcal{D}$.

Proof. Consider (ii): If $\mathcal{S}$ is tight, then every sequence $\{F^\nu\}_{\nu\in\mathbb{N}} \subset \mathcal{S}$ is also tight. Consequently, in view of Proposition 3.5(ii), $\mathcal{D}$ contains the limit points of $\mathcal{S}$. By Proposition 3.5(i), every sequence $\{F^\nu\}_{\nu\in\mathbb{N}} \subset \mathcal{S}$ contains a convergent subsequence, with limit in $\mathrm{cl}\,\mathcal{S}$ and the conclusion follows. Next, consider (i). We first establish the tightness of every convergent sequence in $\mathcal{S}$. Let $\varepsilon > 0$ and $d\!l^{\mathrm{h}}(F^\nu, F) \to 0$, with $F^\nu, F \in \mathcal{S}$. Suppose that P^ν and P are the probability measures corresponding to F^ν and F, respectively. Select a bounded and open set $A_0 \subset \mathbb{R}^m$ such that $P(A_0) \ge 1 - \varepsilon/2$. By Theorem 3.3, P^ν converges weakly to P. It is well known that this implies that $\liminf P^\nu(A_0) \ge P(A_0)$. Consequently, there exists a $\bar\nu$ such that $P^\nu(A_0) \ge P(A_0) - \varepsilon/2$ for all $\nu \ge \bar\nu$. Thus, $P^\nu(A_0) \ge 1 - \varepsilon$ for all $\nu \ge \bar\nu$. Let $\bar A \supset A_0$ be a rectangle. Then,

$$P^\nu(\bar A) \ge P^\nu(A_0) \ge 1 - \varepsilon \text{ for all } \nu \ge \bar\nu.$$

For $\nu = 1, 2, ..., \bar\nu - 1$, select bounded sets $A^\nu \subset \mathbb{R}^m$ such that $P^\nu(A^\nu) \ge 1 - \varepsilon$. We can then find a rectangle $A \supset \bar A \cup (\cup_{\nu=1}^{\bar\nu-1} A^\nu)$. Thus, $P^\nu(A) \ge 1 - \varepsilon$ and $\Delta_A F^\nu \ge 1 - \varepsilon$ for all ν. Since ε was arbitrary, this implies that $\{F^\nu\}_{\nu\in\mathbb{N}}$ is tight. Second, we establish tightness of the whole of $\mathcal{S}$ by means of a contradiction. Suppose that $\mathcal{S}$ is not tight. Then, there exist $1 > \varepsilon > 0$, $\{F^\nu\}_{\nu\in\mathbb{N}} \subset \mathcal{S}$, and rectangles $\{A^\nu\}_{\nu\in\mathbb{N}}$ in $\mathbb{R}^m$ such that $\nu\mathbb{B} \subset A^\nu$ and $\Delta_{A^\nu} F^\nu < 1 - \varepsilon$. Since $\{F^\nu\}_{\nu\in\mathbb{N}} \subset \mathcal{S}$ and $\mathcal{S}$ is compact, there exists a subsequence $\{F^{\nu_k}\}$ and $F \in \mathcal{S}$ such that $F^{\nu_k} \to F$ and, thus $\{F^{\nu_k}\}$ is tight by the virtue of being a convergent sequence. However, by construction, $\{F^{\nu_k}\}$ is not tight and we have reached a contradiction. □

We note that although a ball $\mathbb{B}(F, r) \subset \text{usc-fcns}(\mathbb{R}^m)$, with $F \in \mathcal{D}$, is compact, the subset $\mathbb{B}(F, r) \cap \mathcal{D}$ is neither closed nor tight unless $r = 0$. This is easily realized in the following manner. Let $g : \mathbb{R}^m \to [0, 1]$ be defined such that $g(\xi) = \max\{0, F(\xi) - r\}$ and $r > 0$. Clearly, $d\!l^{\mathrm{h}}(g, F) \le r$. Let $A \subset \mathbb{R}^m$ be such that $\Delta_A(F) \ge 1 - r$. Set $c > 0$ such that[8] $A \subset \{\xi \in \mathbb{R}^m \;:\; \xi \le c\mathbf{1}\}$. Construct the

[8]We use the notation $\mathbf{1} = (1, 1, ..., 1) \in \mathbb{R}^m$.

distribution functions $F^\nu : I\!\!R^m \to [0,1]$ such that $F^\nu(\xi) = 1$ if $c\nu\mathbf{1} \leq \xi$ and $F^\nu(\xi) = g(\xi)$ otherwise. Since $F(\xi) \geq 1-r$ for ξ with $F^\nu(\xi) = 1$, we have that $|F^\nu(\xi) - F(\xi)| \leq r$ for all ξ and thus $d\!l^{\mathrm{h}}(F^\nu, F) \leq r$. However, $\{F^\nu\}$ is not tight and does not tend to a distribution function. The only balls contained in $\mathcal{D}$ are those with zero radius, i.e., $I\!\!B(F, 0)$. We observe that a set-up centered on the metric space $(\mathcal{D}, d\!l^{\mathrm{h}})$, instead of (usc-fcns($I\!\!R^m$), $d\!l^{\mathrm{h}}$), is possible, but has the disadvantage that the space is not complete.

We next give estimates of the hypo-distance between two distribution functions.

3.7 Theorem (estimates of hypo-distance) *For distribution functions $F, G \in \mathcal{D}$, we have that for any $\rho \in [1, \infty]$,*

$$\underline{\eta}(\rho)e^{-\rho} \leq d\!l^{\mathrm{h}}(F, G) \leq e^{-\rho} + (1 - e^{-\rho})\overline{\eta}(\rho) \leq \overline{\eta}(\infty),$$

where

$$\overline{\eta}(\rho) = \inf\big\{\eta \geq 0 \ : \ G(\xi + \eta\mathbf{1}/\sqrt{m}) + \eta \geq F(\xi) \text{ and } F(\xi + \eta\mathbf{1}/\sqrt{m}) + \eta \geq G(\xi) \quad \text{for all } \xi \in 2\rho I\!\!B\big\}$$
$$\underline{\eta}(\rho) = \inf\big\{\eta \geq 0 \ : \ G(\xi + \eta\mathbf{1}) + \eta \geq F(\xi) \text{ and } F(\xi + \eta\mathbf{1}) + \eta \geq G(\xi) \quad \text{for all } \xi \in \rho I\!\!B\big\}.$$

Proof. As the proof is somewhat involved and relies on notation established later, we postpone it to the appendix. □

We note that these estimates are related but not identical to the Levy metric in the one-dimensional case ($m = 1$).

Example 6: Hypo-distance for Exponential Distribution. The previous results provide the following bound on the hypo-distance between the distribution functions F and G given by $F(\xi) = 0$ for $\xi < 0$ and $F(x) = 1 - \exp(-\lambda\xi)$ for $\xi \geq 0$ (the exponential density) and $G(\xi) = 0$ for $\xi < 0$ and $G(\xi) = 1$ for $\xi \geq 0$. In this case, $\overline{\eta}(\rho)$ and $\underline{\eta}(\rho)$ are identical and equal to the Levy metric between F and G for any $\rho \geq 1$. In fact, the value does not change for $\rho \geq 1$. One can then find that $\overline{\eta}(\rho)$ is the root of $\exp(-\lambda\eta) - \eta = 0$. For $\lambda = 1, 10, 100$, and 1000, the roots are 0.5671, 0.1746, 0.0339, and 0.0052, respectively. Thus, since these are independent of $\rho \geq 1$, the upper bounds on the hypo-distances are simply these roots. The lower bound is scaled with $\exp(-1)$.

An ambiguity set can be made "smaller" by including moment restrictions as illustrated next.

Example 7: Ambiguity Sets under Moment Information. Given scalars $\mu_1 < \mu_2$ and $s > 0$, the ambiguity set of distribution functions on $I\!\!R$ given by

$$\mathcal{F}_{\mu_1,\mu_2,s} = \{F \in \mathcal{D} \ : \ F \text{ has mean in } [\mu_1, \mu_2] \text{ and standard deviation in } [0, s]\}$$

is tight and for every $r > 0$ there exist an integer ν_r and $F^1, ..., F^{\nu_r}$ such that

$$\{F^1, ..., F^{\nu_r}\} \subset \mathcal{F}_{\mu_1,\mu_2,s} \subset \bigcup_{\nu=1}^{\nu_r} I\!\!B(F^\nu, r).$$

To establish this fact, let $\varepsilon > 0$ and $\boldsymbol{\xi}$ be a random variable with mean $\mu \in [\mu_1, \mu_2]$ and standard deviation $\sigma \in (0, s]$. Then, Chebyshev's inequality implies that

$$\mathrm{prob}(\mu_1 - s/\sqrt{\varepsilon} < \boldsymbol{\xi} \leq \mu_2 + s/\sqrt{\varepsilon} + 1) \geq 1 - \mathrm{prob}(|\boldsymbol{\xi} - \mu| \geq \sigma/\sqrt{\varepsilon}) \geq 1 - \varepsilon.$$

The special case with zero standard deviation is trivial and, thus, $\mathcal{F}_{\mu_1,\mu_2,s}$ is tight. Proposition 3.6 establishes that $\mathrm{cl}\,\mathcal{F}_{\mu_1,\mu_2,s}$ is compact and is contained in $\mathcal{D}$. The conclusion is then a direct consequence of compactness. Here we see that knowledge about the range of values for the two first moments of the distributions effectively reduces the "ambiguity" to a finite set of distribution functions, at least approximately. Thus, moment information is actually rather powerful.

We end the section by highlighting key properties of expectation functionals due to their frequent appearance in applications.

3.8 Proposition (expectation functionals) *Given a measurable function $\psi : \mathbb{R}^m \to \mathbb{R}$ and $\mathcal{F} \subset \mathcal{D}$, with $\int |\psi(\xi)| dF(\xi) < \infty$ for all $F \in \mathcal{F}$, the functional $\varphi : \mathcal{F} \to \mathbb{R}$ given by*

$$\varphi(F) = \int \psi(\xi) dF(\xi),\ F \in \mathcal{F},$$

is well-defined. Moreover,

(i) φ is convex, in fact linear, whenever $\mathcal{F}$ is a convex set;

(ii) φ is lower semicontinuous at $F \in \mathcal{F}$ whenever the sets of discontinuity of ψ and F fail to intersect and $\psi \geq \beta \in \mathbb{R}$; and

(iii) φ is continuous at $F \in \mathcal{F}$ whenever the sets of discontinuity of ψ and F fail to intersect and for every $F^\nu \to F$, $F^\nu \in \mathcal{F}$,

$$\lim_{\alpha\to\infty} \sup_{\nu\in N} \int_{\{\xi:|\psi(\xi)|\geq\alpha\}} |\psi(\xi)| dF^\nu(\xi) = 0.$$

Proof. Certainly, φ is well-defined in view of the integrability of ψ. Part (i) is obvious from the definition of convexity. For Part (ii), let $F^\nu \to F$, with $F^\nu, F \in \mathcal{F}$, which also implies that F^ν converges weakly to F by Theorem 3.3. Let $\boldsymbol{\xi}^\nu$ and $\boldsymbol{\xi}$ be random vectors distributed according to F^ν and F, respectively. The classical mapping theorem then ensures that $\psi(\boldsymbol{\xi}^\nu)$ converges in distribution to $\psi(\boldsymbol{\xi})$. A standard application of Fatou's lemma establishes the conclusion; see [8, Theorem 25.11]. For Part (iii), the stated uniform integrability assumption ensures that convergence in distribution suffices for convergence in expectation; see for example [8, Theorem 25.12]. $\square$

4 Lopsided Convergence

We set out to apply our recent extension of the variational theory of lopsided convergence [33] for the first time in the context of optimization under stochastic ambiguity, where it appears especially well suited. In this section, we first recall definitions and essential facts about lopsided convergence pertaining to the present context; see [33] for details and proofs. Second, we develop results for optimization under stochastic ambiguity, including sufficient conditions for lopsided convergence.

Associated with (SAP), we consider a family of approximate problems

$$(\mathrm{SAP})^\nu: \quad \min_{x\in X^\nu} \sup_{F\in\mathcal{F}^\nu(x)} \varphi^\nu(x, F),\ \text{where } \mathcal{F}^\nu : X^\nu \rightrightarrows \mathcal{D} \text{ has } \mathcal{F}^\nu(x) \neq \emptyset \text{ for all } x \in X^\nu,$$

where the sets $X^\nu \subset I\!R^n$ are nonempty and the approximate objective functions

$$\varphi^\nu : Z^\nu \to I\!R, \text{ with } Z^\nu := \Big\{(x,F) \in X^\nu \times \mathcal{D} \; : \; F \in \mathcal{F}^\nu(x)\Big\}.$$

The approximating problems $\{(\mathrm{SAP})^\nu\}_{\nu\in N}$ represent a multitude of situations arising in applications that demand "approximations" for computational and/or modeling reasons. We permit approximation in all components of (SAP): the objective function, the region of feasible decision vectors, and the ambiguity set. To concisely represents the three components of $(\mathrm{SAP})^\nu$, we often write $\varphi^\nu_{X^\nu\mathcal{F}^\nu}$. Lopsided convergence as defined next deals with a notion of approximation by the bifunctions $\varphi^\nu_{X^\nu\mathcal{F}^\nu}$ defining $(\mathrm{SAP})^\nu$ of the bifunction $\varphi_{X\mathcal{F}}$ defining (SAP). This convergence in turn ensures convergence of minsup-points and minsup-values of $(\mathrm{SAP})^\nu$ to those of (SAP) as we develop in this section.

4.1 Definitions and Consequences

We start with the definition of lopsided convergence, which in additional to its "basic" form comes in two strengthened forms referred to as *ancillary-tight lop-convergence* and *tight lop-convergence*. Since we use the results of this subsection both for the general (SAP), which involves "outer" decision vector $x \in I\!R^n$ and "inner" variable $F \in \mathcal{D} \subset$ usc-fcns$(I\!R^m)$, as well as for special cases where the inner maximization is over a finite-dimensional space, we simply state results for an inner variable y in a general metric space $(\mathcal{Y}, d_\mathcal{Y})$. Consequently, for a nonempty set $X \subset I\!R^n$ and a set-valued mapping $Y : X \rightrightarrows \mathcal{Y}$, with $Y(x) \neq \emptyset$ for all $x \in X$, we consider a bifunction f_{XY} that is finite-valued for every $x \in X$ and $y \in Y(x)$. Approximating bifunctions $f^\nu_{X^\nu Y^\nu}$ are given similarly.

4.1 Definition (lopsided convergence) *The bifunctions $\{f^\nu_{X^\nu Y^\nu}\}_{\nu\in N}$ lop-converge to f_{XY} if*

(i) for all $x^\nu \to x \in X$, with $x^\nu \in X^\nu$, and $y \in Y(x)$, there exist $y^\nu \to y$, with $y^\nu \in Y^\nu(x^\nu)$, such that $\liminf f^\nu(x^\nu, y^\nu) \geq f(x,y)$;

(ii) for all $x^\nu \to x \notin X$, with $x^\nu \in X^\nu$, there exist $y^\nu \in Y^\nu(x^\nu)$, such that $f^\nu(x^\nu, y^\nu) \to \infty$;

(iii) for all $x \in X$, there exists $x^\nu \to x$, with $x^\nu \in X^\nu$, such that for all $y^\nu \to y \in \mathcal{Y}$, with $y^\nu \in Y^\nu(x^\nu)$,

$$\begin{aligned} &\limsup f^\nu(x^\nu, y^\nu) \leq f(x,y) && \text{if } y \in Y(x);\\ &f^\nu(x^\nu, y^\nu) \to -\infty && \text{otherwise.}\end{aligned}$$

The convergence is ancillary tight if, in addition, for every $\varepsilon > 0$ and $x^\nu \to x$ selected in (iii), there exists a compact set $B_\varepsilon \subset \mathcal{Y}$ and an integer ν_ε such that

$$\sup_{y\in Y^\nu(x^\nu)\cap B_\varepsilon} f^\nu(x^\nu, y) \geq \sup_{y\in Y^\nu(x^\nu)} f^\nu(x^\nu, y) - \varepsilon \text{ for all } \nu \geq \nu_\varepsilon.$$

The convergence is tight if, in addition to all the above, for any $\varepsilon > 0$ there exists a compact set $A_\varepsilon \subset I\!R^n$ and an integer ν_ε such that

$$\inf_{x\in X^\nu\cap A_\varepsilon} \sup_{y\in Y^\nu(x)} f^\nu(x,y) \leq \inf_{x\in X^\nu} \sup_{y\in Y^\nu(x)} f^\nu(x,y) + \varepsilon \text{ for all } \nu \geq \nu_\varepsilon.$$

The requirements of ancillary-tight and tight lop-convergence can be viewed as relaxed "uniform" compactness assumptions on $Y^\nu(x)$ and X^ν. Obviously, ancillary-tightness is satisfied if all $Y^\nu(x^\nu)$ are contained in a compact set. Tightness is ensured if in addition all X^ν are contained in a compact set. In both cases, many other situations without compactness will also satisfy the requirements.

Next, we turn to the implications of lopsided convergence and denote by

$$\varepsilon\text{-}\operatorname{argmin}_X \sup_Y f := \Big\{x \in X \ : \ \sup_{y\in Y(x)} f(x,y) \leq \inf_X \sup_Y f + \varepsilon\Big\}$$

a set of ε-optimal solutions, $\varepsilon \geq 0$. If $\varepsilon = 0$, the set consists of the minsup-points of the bifunction f_{XY}.

4.2 Proposition (consequences of lop-convergence) *Suppose that $\{f^\nu_{X^\nu Y^\nu}\}_{\nu\in N}$ lop-converge to f_{XY} and $\sup_{y\in Y(x)} f(x,y) < \infty$ for some $x \in X$. Then the following hold:*

(i) If there is a sequence $\{x^\nu\}_{\nu\in N}$, with $x^\nu \in \operatorname{argmin}_{X^\nu} \sup_{Y^\nu} f^\nu$, that has a cluster point, then

$$\operatorname{liminf}\,(\inf_{X^\nu} \sup_{Y^\nu} f^\nu) \geq \inf_X \sup_Y f.$$

(ii) If the convergence is ancillary tight, then

$$\operatorname{limsup}\ (\inf_{X^\nu} \sup_{Y^\nu} f^\nu) \leq \inf_X \sup_Y f.$$

(iii) If the convergence is ancillary tight and there is a sequence $x^k \to x$, with $x^k \in \operatorname{argmin}_{X^{\nu_k}} \sup_{Y^{\nu_k}} f^{\nu_k}$ for some increasing subsequence $\{\nu_1,\nu_2,...\} \subset I\!N$, then

$$x \in \operatorname{argmin}_X \sup_Y f \quad \textit{and} \quad \lim_{k\to\infty} (\inf_{X^{\nu_k}} \sup_{Y^{\nu_k}} f^{\nu_k}) = \inf_X \sup_Y f.$$

(iv) If the convergence is tight and $\inf_X \sup_Y f$ is finite, then

$$\inf_{X^\nu} \sup_{Y^\nu} f^\nu \to \inf_X \sup_Y f.$$

Moreover, for every $x^* \in \operatorname{argmin}_X \sup_Y f$, there exist an infinite subsequence N of $I\!N$, $\varepsilon^\nu \searrow 0$, and $\{x^\nu\}_{\nu\in N}$ such that

$$x^\nu \in \varepsilon^\nu\text{-}\operatorname{argmin}_{X^\nu} \sup_{Y^\nu} f^\nu \quad \textit{and} \quad x^\nu \to^N x^*.$$

Conversely, if such N, $\{\varepsilon^\nu\}$ and $\{x^\nu\}$ exist and $x^\nu \to^N x^{**}$, then

$$\inf_{X^\nu} \sup_{Y^\nu} f^\nu \to^N \sup_{y\in Y(x^{**})} f(x^{**},y) = \inf_X \sup_Y f.$$

It is apparent from this proposition, which is a collection of results from [33], that lopsided convergence becomes a central property in the study of approximations of (SAP).

It is well-known that the infimum over a compact set of an extended real-valued lower semicontinuous (lsc) function is attained. Consequently, if $\sup_{y\in Y(\cdot)} f(\cdot,y)$ is lsc on a compact set X, then there exists a minsup-point of f_{XY}. We next state an existence result for minsup-points that relaxes the compactness requirement but instead imposes them on approximating problems. The result compiles Proposition 3.6 and Theorem 3.15 in [33].

4.3 Proposition (existence of minsup-point) *Suppose that the bifunctions* $\{f^\nu_{X^\nu Y^\nu}\}_{\nu\in N}$ *lop-converge ancillary-tightly to* f_{XY}, X^ν *is compact, and* $\sup_{y\in Y(x)} f(x,y) < \infty$ *for some* $x \in X$. *Moreover, suppose that for each* ν, *the following holds:*

(i) Y^ν *is inner semicontinuous*[9].

(ii) For every $x^k \to x \in X^\nu$, *with* $x^k \in X^\nu$, *and* $y^k \to y \in Y^\nu(x)$, *with* $y^k \in Y^\nu(x^k)$, $\liminf f^\nu(x^k, y^k) \geq f^\nu(x,y)$.

Then, for all ν *there exists a minsup-point* x^ν *of* $f^\nu_{X^\nu Y^\nu}$ *and every cluster point of* $\{x^\nu\}_{\nu\in N}$ *is a minsup-point of* f_{XY}.

If Y^ν *is constant on* X^ν *for all* ν, *then conditions (i) and (ii) can be replaced by the requirement that* $f^\nu(\cdot, y)$ *is lsc for all* ν *and* $y \in Y^\nu(x)$, $x \in X^\nu$.

Proposition 4.3 does not ensure the existence of a cluster point of $\{x^\nu\}_{\nu\in N}$. Still, the proposition provides an approach for establishing the existence of a minsup-point of (SAP) without requiring compactness of X and $\mathcal{F}(x)$ for $x \in X$. One can first construct a sequence $\{\varphi^\nu_{X^\nu \mathcal{F}^\nu}\}_{\nu\in N}$, with the required properties, that lop-converges to $\varphi_{X\mathcal{F}}$ and, second, prove that $\{x^\nu\}_{\nu\in N}$ has a cluster point.

4.2 Lop-Convergence in Optimization under Stochastic Ambiguity

Relying on the previous subsection, we next develop a series of specific results under various assumptions about (SAP) and (SAP)$^\nu$. We start with a sufficient condition for ancillary-tight and tight lop-convergence.

4.4 Proposition (sufficient condition for tight lop-convergence) *Suppose that* $\{\varphi^\nu_{X^\nu \mathcal{F}^\nu}\}_{\nu\in N}$ *lop-converges to* $\varphi_{X\mathcal{F}}$.

(i) The convergence is ancillary-tight if for every sequence $x^\nu \to x$ *selected in Definition 4.1(iii) to establish lop-convergence,* $\{\mathcal{F}^\nu(x^\nu)\}_{\nu\in N}$ *is tight.*

(ii) The convergence is tight, if it is ancillary-tight and there exists a bounded set $A \subset I\!R^n$ *such that* $\{X^\nu\}_{\nu\in N} \subset A$.

Proof. In view of Proposition 3.6, $\mathrm{cl}\{\mathcal{F}^\nu(x^\nu)\}_{\nu\in N}$ is compact and therefore furnishes the compact set B_ε required in Definition 4.1. Thus, assumption (i) ensures ancillary-tightness. With the addition of assumption in item (ii), tight lop-convergence is guaranteed by Definition 4.1. □

Applications give rise to numerous approximations of $\mathcal{F}(x)$ by some set $\mathcal{F}^\nu(x)$. It is worth recording the following special case involving empirical distribution functions, which of course converge weakly (and in stronger senses too) to the underlying distribution function.

4.5 Proposition (ancillary-tight lop-convergence under sampling approximations) *Suppose that* $\mathcal{F} = \{F^0\} \subset \mathcal{D}$ *and* $\{\boldsymbol{\xi}^\nu\}_{\nu\in N}$ *is a sequence of independent and identically distributed random vectors with*

[9] A set-valued mapping $S : X \rightrightarrows \mathcal{Y}$ is inner semicontinuous if for every $x^k \to x \in X, x^k \in X$, $\liminf S(x^k) \supset S(x)$.

values in $I\!R^m$ that is defined on a probability space $(\Omega, \mathcal{A}, \mu)$ and F^0 is the distribution function of $\boldsymbol{\xi}^1$. Let

$$F^\nu(\cdot, \omega) = \frac{1}{\nu} \sum_{j=1}^{\nu} I(\xi^j(\omega) \leq \cdot)$$

be the corresponding empirical distribution functions and $\mathcal{F}_\omega^\nu = \{F \in \mathcal{D} \;:\; d\!l^{\rm h}(F, F^\nu(\cdot, \omega)) \leq \varepsilon^\nu\}$, with $\varepsilon^\nu \geq 0$.

If X is closed, $\varepsilon^\nu \to 0$, and for all $x^\nu \to x \in X$, with $x^\nu \in X$, and $G^\nu \to F^0$, with $G^\nu \in \mathcal{D}$,

$$\operatorname{liminf} \varphi^\nu(x^\nu, G^\nu) \geq \varphi(x, F^0) \text{ and } \operatorname{limsup} \varphi^\nu(x, G^\nu) \leq \varphi(x, F^0),$$

then $\varphi^\nu_{X\mathcal{F}^\nu_\omega}$ lop-converges ancillary-tightly to $\varphi_{X\mathcal{F}}$ for μ-almost every $\omega \in \Omega$.

Proof. As discussed in Example 5, $F^\nu(\cdot, \omega) \to F^0$ for μ-almost every $\omega \in \Omega$. Let ω be such that this convergence takes place. We first establish lop-convergence. Let $x^\nu \to x \in X$, with $x^\nu \in X$. Since $\operatorname{liminf} \varphi^\nu(x^\nu, F^\nu(\cdot, \omega)) \geq \varphi(x, F^0)$, Definition 4.1(i) holds. Definition 4.1(ii) does not apply due to the closedness of X. For Definition 4.1(iii) take $x^\nu = x \in X$. Let $G^\nu \to F \in$ usc-fcns($I\!R^m$), with $G^\nu \in \mathcal{F}_\omega^\nu$. Since $F^\nu(\cdot, \omega) \to F^0$ and $\varepsilon^\nu \to 0$, $F = F^0$. Definition 4.1(iii) then follows from the fact that $\operatorname{limsup} \varphi^\nu(x, G^\nu) \leq \varphi(x, F^0)$. Thus, lop-converges is established. Second, for $\varepsilon > 0$, define $B_\varepsilon = I\!B(F^0, \varepsilon)$, which is compact. There exists a ν_ε such that $d\!l^{\rm h}(F^\nu(\cdot, \omega), F^0) \leq \varepsilon/2$ and $\varepsilon^\nu \leq \varepsilon/2$ for all $\nu \geq \nu_\varepsilon$. Hence, $\mathcal{F}_\omega^\nu \subset B_\varepsilon$ for all $\nu \geq \nu_\varepsilon$. In view of Definition 4.1, the convergence is therefore ancillary tight. □

The proposition shows the practically useful result that if the feasible region X is not approximated and the objective function is approximated in some "continuous" way, then not only the empirical distribution functions but also an approximation based on a "robust band" around them lead to lopsided convergence and thus convergence of minsup-points and minsup-values. This provides a justification for the strategy of "robustify" an empirical distribution function that is feared to deviate substantially from the actual distribution function.

Specific instances of Example 2 are those involving (first-order) superquantile risk measures. These provide an opportunity to illustrate the above concepts in an elementary manner.

Example 9: Superquantile Risk Minimization. Considering a finite probability space, we let a reference probability measure P_0 satisfy $\sum_{j=1}^J P_0(\xi^j) = 1$ for some $\xi^1, ..., \xi^J \in I\!R^m$. For $\alpha \in [0, 1)$, the problem of minimizing by choice of $x \in X$ the α-superquantile risk of $\psi(x, \boldsymbol{\xi})$, with $\boldsymbol{\xi}$ distributed according to P_0, then takes the form

$$\min_{x \in X} \sup_{F \in \mathcal{F}} \int \psi(x, \xi) dF(\xi), \quad \text{where } \mathcal{F} = \Bigg\{ F \in \mathcal{D} \;:\; \text{there exists } P \in \mathcal{M} \text{ with } P(\xi^j) \leq \frac{P_0(\xi^j)}{1-\alpha},$$

$$j = 1, ..., J, \text{ and } \sum_{j=1}^J P(\xi^j) = 1 \text{ such that } F(\xi) = P(S_\xi) \text{ for all } \xi \in I\!R^m \Bigg\};$$

see for example [28]. It is obvious that this problem is equivalent to

$$\min_{x \in X} \max_{y \in Y_\alpha} \sum_{j=1}^J \psi(x, \xi^j) P_0(\xi^j) y_j, \text{ where } Y_\alpha := \left\{ y \in I\!R^J \;:\; 0 \leq y_j \leq \frac{1}{1-\alpha} \text{ for all } j, \; \sum_{j=1}^J P_0(\xi^j) y_j = 1 \right\}.$$

Since α can be interpreted as the level of risk-averseness, it is of interest to examine the effect of changing α; see [24] for a detailed study of this subject. Let $f(x,y) = \sum_{j=1}^{J} \psi(x,\xi^j)P_0(\xi^j)y_j$ be the objective function, which we do not approximate, and let $\psi(\cdot,\xi^j)$ be finite and lsc on a nonempty closed set X for all $j = 1,...,J$. Then, it easy to establish that $f_{XY_{\alpha^\nu}}$ lop-converges to f_{XY_α} directly from Definition 4.1 when $\alpha^\nu \to \alpha$. Specifically, suppose that $x^\nu \to x \in X$, with $x^\nu \in X$, and $y \in Y_\alpha$. Obviously, there exists $y^\nu \in Y_{\alpha^\nu}$ such that $y^\nu \to y$. Since $\psi(\cdot,\xi^j)$ is lsc, liminf $f(x^\nu,y^\nu) \geq f(x,y)$ and Definition 4.1(i) is established. Item (ii) in that definition is automatic because X is closed. It only remains to show item (iii). Let $x \in X$ and select $x^\nu = x$. Moreover, let $y^\nu \to y \in I\!R^J$, with $y^\nu \in Y_{\alpha^\nu}$. Since Y_α is compact, y must necessarily be in that set when it is a limit of points in Y_{α^ν}. Since $f(x,y^\nu) \to f(x,y)$, item (iii) also holds. Thus, $f_{XY_{\alpha^\nu}}$ lop-converges to f_{XY_α}. Since $\{Y_{\alpha^\nu}\}_{\nu\in N}$ is contained in a compact set, the convergence is ancillary tight.

A simple example from investment planning illustrates the concepts in the context of robust optimization.

Example 10: Robust Investment Problem. We consider a specific case taken from [6, 41]. Suppose that $m = n$, $\psi(x,\xi) = \langle -\xi, x\rangle$ is the negative of the return of an investment portfolio under allocation x and return ξ, and the uncertainty set

$$\Xi_\theta(x) = \left\{ \xi \in I\!R^n \ : \ \sum_{i=1}^{n} \left(\frac{\xi_i - \bar{\xi}_i}{\sigma_i}\right)^2 \leq b_\theta(x) \right\}, \text{ with } \sigma_i \in (0,\infty) \text{ and } \bar{\xi}_i \in I\!R,$$

where the "budget of uncertainty" $b_\theta(x)$ might be independent of x or, as in [41, p.17], become larger as x_i deviates from the equal allocation $1/n$. Specifically, let

$$b_\theta(x) = \theta\left(1 + \sum_{i=1}^{n}(x_i - 1/n)^2\right), \text{ for } \theta \in [0,\infty).$$

It is interesting to examine the effect of varying the "nominal budget of uncertainty" θ. If X is closed and $\theta^\nu \to \theta$, $\theta^\nu \geq 0$, it is easy to establish that $\psi_{X\Xi_{\theta^\nu}}$ lop-converges to $\psi_{X\Xi_\theta}$ using the definition of lop-convergence. The convergence is actually ancillary tight since is suffices to consider $x^\nu = x \in X$ in Definition 4.1(iii) and $\{\Xi_{\theta^\nu}(x)\}_{\nu\in N}$ is contained in a compact set for all $x \in X$. Consequently, in view of Proposition 4.2, minsup-points and minsup-values are "stable" in some sense under changes in the nominal budget of uncertainty.

An example with a tilt towards computational methods for solving difficult problems involving decision-dependent uncertainty sets is the next subject.

Example 11: Robust Optimization and Generalized Semi-infinite Programming. A class of robust optimization problems are the generalized semi-infinite programs

$$\min_{x\in X} \sup_{\xi\in\Xi(x)} f(x,\xi), \text{ with } \Xi(x) = \{\xi \in \Xi_0 \subset I\!R^m \ : \ g(x,\xi) \leq 0\} \text{ and } g : I\!R^n \times I\!R^m \to I\!R^k;$$

see [41] for a general treatment. These problems can be solved approximately by considering the approximate problems

$$\min_{x\in X} \sup_{\xi\in\Xi_0} f^\nu(x,\xi) = f(x,\xi) - t^\nu g(x,\xi)_+,$$

where $v_+ = \max\{0, v_1, v_2, ..., v_k\}$ for any vector $v = (v_1, ..., v_k)$ and $t^\nu \to \infty$ are positive penalties; see [31, 30] for algorithms along these lines. Suppose that X and Ξ_0 are closed, f and g are continuous on $X \times \Xi_0$, and Ξ is a nonempty-valued and continuous set-valued mapping on X. Lop-convergence of $f^\nu_{X\Xi^0}$ to $f_{X\Xi}$ can be established by means of Definition 4.1 as follows. We note that this is a case where the uncertainty set $\Xi(x)$ is not approximated directly, but still lop-convergence can be established. First we consider Definition 4.1(i) and let $x^\nu \to x \in X$, $x^\nu \in X$, and $\xi \in \Xi(x)$. Select $\{\xi^\nu\}_{\nu \in N} \subset \Xi_0$ such that $\xi^\nu \to \xi$ and $g(x^\nu, \xi^\nu) \leq 0$ for all ν. Such a sequence obviously exists in view of the continuity assumption about Ξ. Thus,

$$f^\nu(x^\nu, \xi^\nu) = f(x^\nu, \xi^\nu) - t^\nu g(x^\nu, \xi^\nu)_+ = f(x^\nu, \xi^\nu) \to f(x, \xi),$$

which established the first requirement. Definition 4.1(ii) is automatically satisfied since X is closed. It only remains to consider Definition 4.1(iii). Let $x \in X$ and take $x^\nu = x$ for all ν. Moreover, let $\xi^\nu \to \xi$ be a sequence in Ξ_0. If $g(x, \xi) \leq 0$, then

$$\limsup f^\nu(x^\nu, \xi^\nu) = \limsup f(x, \xi^\nu) - t^\nu g(x, \xi^\nu)_+ \leq f(x, \xi).$$

If $g(x, \xi) > 0$, then

$$f^\nu(x^\nu, \xi^\nu) = f(x, \xi^\nu) - t^\nu g(x, \xi^\nu)_+ \to -\infty$$

because $t^\nu \to \infty$ and f, g are continuous. Since Ξ_0 is closed, this establishes lopsided convergence. Thus, through Proposition 4.2, this provides justification for algorithms based on the solution of the approximate problem.

In stochastic optimization, the probability distribution function F of the random vector might depend of the decision x. Even in the absence of ambiguity, such problems are considered extremely hard to solve. The present framework facilitates the development of approximation schemes as follows.

Example 12: Decision-Dependent Distributions in Stochastic Optimization. The problem $\min_{x \in X} \varphi(x, F^0(\cdot; x))$ involving an x-dependent distribution function $F^0(\cdot; x) \in \mathcal{D}$ can be reformulated as

$$\min_{x \in X} \sup_{F \in \mathcal{F}(x)} \varphi(x, F), \text{ with } \mathcal{F}(x) = \{F^0(\cdot; x)\} \subset \mathcal{D} \text{ a singleton.}$$

Conceptually, a possible approach to solving this problem is to consider the approximation

$$\min_{x \in X} \sup_{F \in \mathcal{D}} \varphi^\nu(x, F) = \varphi(x, F) - t^\nu d\!l^{\mathrm{h}}(F, F^0(\cdot; x)), \text{ for } t^\nu > 0.$$

If the distribution function $F^0(\cdot; x)$ varies continuously in x, i.e., $d\!l^{\mathrm{h}}(F^0(\cdot; x^\nu), F^0(\cdot; x)) \to 0$ whenever $x^\nu \to x$, and φ is continuous, then an argument following the pattern of the previous example establishes that the approximate problems lop-converge to the original one as long as $t^\nu \to \infty$. This fact provides new possibilities of algorithmic development for this challenging problem.

5 Quantification of Lop-Convergence and Solution Estimates

We establish in [33] that lopsided convergence is quantified by the lop-distance, which then can be used to estimate rate of convergence for optimization problems with stochastic ambiguity. This section first summarizes the key definitions and results regarding the lop-distance. Second, we provide for the first time estimates of the lop-distance. Again, as in Subsection 4.1, we consider a general metric space $(\mathcal{Y}, d_{\mathcal{Y}})$ instead of (usc-fcns$(\mathbb{R}^m), d\!l^{\mathrm{h}})$ as the space for the inner maximization. This extension comes with no additional complication. We show that the lop-distance between two bifunctions provides bounds on the difference between the corresponding minsup-points and minsup-values. Third, we provide specific illustrations in the context of robust and risk-averse optimization.

5.1 Lop-Distance

For a bifunction f_{XY}, with $X \subset \mathbb{R}^n$ nonempty, $Y : \mathbb{R} \rightrightarrows \mathcal{Y}$, $Y(x) \neq \emptyset$ for all $x \in X$, and $f(x, y)$ finite for all $x \in X$ and $y \in Y(x)$, we define its *sup-projection* as the function $h : X \to \overline{\mathbb{R}}$ given by

$$h(x) := \sup_{y \in Y(x)} f(x, y), \text{ for } x \in X.$$

Since $Y(x) \neq \emptyset$ for all $x \in X$, we have that $h > -\infty$, but the effective domain $\operatorname{dom} h := \{x \in X \,:\, h(x) < \infty\}$ might be strictly contained in X. It is obvious that the minsup-points of f_{XY} are identical to the minimizers of h on X. Thus, sup-projections will be central to the following development. In fact, the key quantity is the *epi-distance* between two sup-projections as defined next.

The epi-graph of a function $g : X \to \overline{\mathbb{R}}$ is denoted by

$$\operatorname{epi} g := \{(x, x_0) \in X \times \mathbb{R} \,:\, g(x) \leq x_0\}.$$

We define, parallel to the hypo-distance $d\!l^{\mathrm{h}}$ on the space usc-fcns$(\mathbb{R}^m)$, the epi-distance $d\!l^{\mathrm{e}}$ on the space lsc-fcns$(\mathbb{R}^n)$ of lsc functions from $\mathbb{R}^n$ to $\overline{\mathbb{R}}$, now excluding the function that is identical to ∞. Specifically, for $g, g' \in$ lsc-fcns$(\mathbb{R}^n)$, let the epi-distance

$$d\!l^{\mathrm{e}}(g, g') := \int_0^\infty d\!l^{\mathrm{e}}_\rho(g, g') e^{-\rho} d\rho,$$

where the *ρ-epi-distance*, $\rho \geq 0$, is given by

$$d\!l^{\mathrm{e}}_\rho(g, g') := \max\left\{ |\operatorname{dist}(\bar{x}, \operatorname{epi} g) - \operatorname{dist}(\bar{x}, \operatorname{epi} g')| \,:\, \|\bar{x}\|_{\mathbb{S}} \leq \rho \right\},$$

where we again use the norm defined in (2) and dist from (3), now on the space $\mathbb{R}^n \times \mathbb{R}$. The parallel with $d\!l^{\mathrm{h}}$ is obvious and, in fact, for $g, g' \in$ lsc-fcns$(\mathbb{R}^n)$, $-g, -g' \in$ usc-fcns$(\mathbb{R}^n)$, and $d\!l^{\mathrm{e}}(g, g') = d\!l^{\mathrm{h}}(-g, -g')$. Consequently, for $g^\nu, g \in$ lsc-fcns$(\mathbb{R}^n)$, g^ν epi-converges to g if and only if $d\!l^{\mathrm{e}}(g^\nu, g) \to 0$. Moreover, (lsc-fcns$(\mathbb{R}^m), d\!l^{\mathrm{e}})$ is a complete separable metric space; see for example [29, Theorem 7.58] and [32, Corrolary 3.6].

With this background, we are ready to state the definition of lop-distance between bifunctions. Since we rely on the epi-distance between the corresponding sup-projections, we limit the scope to bifunctions that has sup-projections in lsc-fcns$(\mathbb{R}^n)$[10]. This is not a strong assumption because, for example, if f_{XY} is lsc as a function of both variables and Y is inner semicontinuous, then its sup-projection is lsc.

[10] Since lsc-fcns$(\mathbb{R}^n)$ consists of extended real-valued functions on $\mathbb{R}^n$ and the sup-projection h of f_{XY} is only defined on X, we here and below abuse notation slightly by letting h also denote the extension of the sup-projection to $\mathbb{R}^n$ by assigning it the value ∞ outside X.

Alternatively, if the set-valued mapping Y is constant on X, then it suffices to have $f(\cdot, y)$ lsc for all $y \in Y(x)$ and some $x \in X$; see Proposition 3.6 in [33].

Our main motivation for defining the lop-distance is to apply it in the study of minsup-points of a bifunctions. Thus, informally, we would like to say that two bifunctions are close if their minsup-points and minsup-values are close; or equivalently that the optimal solutions and optimal values of the corresponding sup-projections are close. As we see below, this is indeed the case if the sup-projections corresponding to the bifunctions are close in the sense of the epi-distance. Consequently, we settled on the following definition of lop-distance in [33].

5.1 Definition (lop-distance) *The lop-distance between two bifunctions f_{XY} and $f'_{X'Y'}$, with corresponding sup-projections h and h' in lsc-fcns($\mathbb{R}^n$), is*

$$d\!l^{1}\big(f_{XY}, f'_{X'Y'}\big) := d\!l^{\mathrm{e}}(h, h').$$

Theorem 4.3 in [33] establishes that if the bifunctions $f^{\nu}_{X^{\nu}Y^{\nu}}$ lop-converge ancillary-tightly to f_{XY} and they all have sup-projections in lsc-fcns($\mathbb{R}^n$), then $d\!l^{1}(f^{\nu}_{X^{\nu}Y^{\nu}}, f_{XY}) \to 0$. The converse also holds in some sense after passing to equivalence classes; see [33] for details. Consequently, the lop-distance provides a central tool in estimating the distance between bifunctions and their minsup-points and minsup-values. Next, for the first time we set out to estimate the lop-distance.

5.2 Estimates of Lop-Distance

It is apparent that estimates of the lop-distance between two bifunctions are intimately connected with estimates of the epi-distance between their sup-projections. We provide one estimate, show its implication for the distance between minsup-points and minsup-values, and end with further details about the estimate under additional assumptions.

For $g, g' \in$ lsc-fcns($\mathbb{R}^n$), we the define the auxiliary quantity

$$\hat{d\!l}^{\mathrm{e}}_{\rho}(g, g') := \inf\Big\{\eta \geq 0 \ : \ \operatorname{epi} g \cap \rho\mathbb{S} \subset \operatorname{epi} g' + \eta\mathbb{S}, \ \operatorname{epi} g' \cap \rho\mathbb{S} \subset \operatorname{epi} g + \eta\mathbb{S}\Big\},$$

where $\mathbb{S}$ is the unit ball at the origin of $\mathbb{R}^n \times \mathbb{R}$ under the norm $\|\cdot\|_{\mathbb{S}}$; see (2). This quantity is usually more accessible than $d\!l^{\mathrm{e}}_{\rho}$ as seen below. As applied to the sup-projections of the bifunctions of interest, it provides a key estimate.

5.2 Theorem (estimates of lop-distance) *For bifunctions $f_{XY}, f'_{X'Y'}$, with sup-projections $h, h' \in$ lsc-fcns($\mathbb{R}^n$), we have for any $\rho \geq 0$,*

$$\begin{aligned}(1 - e^{-\rho})|d - d'| + e^{-\rho}\hat{d\!l}^{\mathrm{e}}_{\rho}(h, h') &\leq d\!l^{1}\big(f_{XY}, f'_{X'Y'}\big)\\ &\leq (1 - e^{-\rho})\hat{d\!l}^{\mathrm{e}}_{\bar{\rho}}(h, h') + e^{-\rho}(\max\{d, d'\} + \rho + 1),\end{aligned}$$

where $d = \operatorname{dist}(0, \operatorname{epi} h)$, $d' = \operatorname{dist}(0, \operatorname{epi} h')$, and $\bar{\rho} \geq 2\rho + \max\{d, d'\}$.

Proof. This result parallels [29, Exercise 7.60], but we still provide a proof as that source omits a direct proof and our setting rely on the norm $\|\cdot\|_{\mathbb{S}}$ (not $\|\cdot\|$). We start by mimicking the arguments in the proof of [29, Lemma 4.41]. Clearly,

$$d\!l^{\mathrm{e}}(h, h') = \int_0^{\rho} d\!l^{\mathrm{e}}_{\tau}(h, h')e^{-\tau}d\tau + \int_{\rho}^{\infty} d\!l^{\mathrm{e}}_{\tau}(h, h')e^{-\tau}d\tau.$$

Since $d\!\!l^{\rm e}_\tau(h,h')$ is nondecreasing as τ increases, we have that

$$d\!\!l^{\rm e}_0(h,h')\int_0^\rho e^{-\tau}d\tau \le \int_0^\rho d\!\!l^{\rm e}_\tau(h,h')e^{-\tau}d\tau \le d\!\!l^{\rm e}_\rho(h,h')\int_0^\rho e^{-\tau}d\tau.$$

and

$$d\!\!l^{\rm e}_\rho(h,h')\int_\rho^\infty e^{-\tau}d\tau \le \int_\rho^\infty d\!\!l^{\rm e}_\tau(h,h')e^{-\tau}d\tau \le \int_\rho^\infty [\max\{d,d'\}+\tau]e^{-\tau}d\tau,$$

where the last inequality follows from the fact that $d\!\!l^{\rm e}_\tau(h,h') \le \max\{d,d'\}+\tau$ by the triangle inequality. Carrying out the integrations on the left- and right-hand sides, we obtain that

$$(1-e^{-\rho})|d-d'| + e^{-\rho}d\!\!l^{\rm e}_\rho(h,h') \le d\!\!l^{\rm e}(h,h') \le (1-e^{-\rho})d\!\!l^{\rm e}_\rho(h,h') + e^{-\rho}(\max\{d,d'\}+\rho+1). \quad (5)$$

Next, we establish the relation between $d\!\!l^{\rm e}_\rho$ and $\hat{d\!\!l}^{\rm e}_\rho$. We following the line of arguments in the proof of [29, Lemma 4.34]. Suppose that C and D are closed subsets of $I\!\!R^n \times I\!\!R$, $\varepsilon > 0$, $\rho > 0$, and $\rho' \ge 2\rho + \operatorname{dist}(0,C)$. We first show that $\operatorname{dist}(\cdot,D) \le \operatorname{dist}(\cdot,C)+\varepsilon$ on $\rho\mathbb{S}$ implies that $C\cap\rho\mathbb{S} \subset D+\varepsilon\mathbb{S}$. The claim is trivial if C is empty. For nonempty C, we have for every $\bar x \in C\cap\rho\mathbb{S}$ that $\operatorname{dist}(\bar x, D) \le \varepsilon$. As D is closed, we have that $C\cap\rho\mathbb{S} \subset D+\varepsilon\mathbb{S}$. Second, we establish that $C\cap\rho'\mathbb{S} \subset D+\varepsilon\mathbb{S}$ implies $\operatorname{dist}(\cdot,D) \le \operatorname{dist}(\cdot,C)+\varepsilon$ on $\rho\mathbb{S}$. For any $\bar x \in I\!\!R^{n+1}$,

$$\begin{aligned}\operatorname{dist}(\bar x, C\cap\rho'\mathbb{S}) &\ge \operatorname{dist}(\bar x, D+\varepsilon\mathbb{S}) = \inf\{\|(\bar y+\varepsilon\bar z)-\bar x\|_{\mathbb{S}} \;:\; \bar y \in D, \bar z\in\mathbb{S}\}\\ &\ge \inf\{\|\bar y-\bar x\|_{\mathbb{S}} - \varepsilon\|\bar z\|_{\mathbb{S}} \;:\; \bar y\in D, \bar z\in\mathbb{S}\} = \operatorname{dist}(\bar x, D)-\varepsilon.\end{aligned}$$

Thus, $\operatorname{dist}(\cdot,D) \le \operatorname{dist}(\cdot, C\cap\rho'\mathbb{S})+\varepsilon$ on $I\!\!R^{n+1}$. It remains to establish that $\operatorname{dist}(\bar x, C\cap\rho'\mathbb{S}) = \operatorname{dist}(\bar x, C)$ when $\bar x\in\rho\mathbb{S}$ and $\rho' \ge 2\rho+\operatorname{dist}(0,C)$. So let $\bar x \in \rho\mathbb{S}$ and $\bar y \in \operatorname{argmin}_{\bar y'\in C}\|\bar x-\bar y'\|_{\mathbb{S}}$, which exists since C is closed. The implication is established if $\bar y \in \rho'\mathbb{S}$. This is indeed the case because $\|\bar y\|_{\mathbb{S}} \le \|\bar x\|_{\mathbb{S}} + \|\bar y-\bar x\|_{\mathbb{S}}$, with $\|\bar y-\bar x\|_{\mathbb{S}} = \operatorname{dist}(\bar x, C) \le \operatorname{dist}(\bar x, 0)+\operatorname{dist}(0,C)$. Consequently,

$$\|\bar y\|_{\mathbb{S}} \le 2\|\bar x\|_{\mathbb{S}} + \operatorname{dist}(0,C) \le 2\rho + \operatorname{dist}(0,C) \le \rho'.$$

Applying these two implications, first with $C = \operatorname{epi} h$ and $D = \operatorname{epi} h'$ and second with $C = \operatorname{epi} h'$ and $D = \operatorname{epi} h$, we obtain that

$$\hat{d\!\!l}^{\rm e}_\rho(h,h') \le d\!\!l^{\rm e}_\rho(h,h') \le \hat{d\!\!l}^{\rm e}_{\bar\rho}(h,h') \text{ for } \bar\rho \ge 2\rho+\max\{d,d'\}.$$

The result is then a combination of these inequalities and (5). □

A convenient alternative formula for $\hat{d\!\!l}^{\rm e}_\rho$ is given next, where we adopt the notation $\operatorname{lev}_\rho g := \{x \in I\!\!R^n \;:\; g(x) \le \rho\}$.

5.3 Proposition (alternative expression for auxillary quantity) *For $g, g' \in$ lsc-fcns$(I\!\!R^n)$ and $\rho \ge 0$,*

$$\begin{aligned}\hat{d\!\!l}^{\rm e}_\rho(g,g') = \inf\Big\{\eta \ge 0 \;:\; &\min_{y\in B(x,\eta)} g'(y) \le \max\{g(x),-\rho\}+\eta, \forall x\in\rho I\!\!B\cap\operatorname{lev}_\rho g\\ &\min_{y\in B(x,\eta)} g(y) \le \max\{g'(x),-\rho\}+\eta, \forall x\in\rho I\!\!B\cap\operatorname{lev}_\rho g'\Big\}.\end{aligned}$$

Proof. Since we are dealing with epi-graphs on both sides, $\eta \geq 0$ satisfies $\operatorname{epi} g \cap \rho\mathbb{S} \subset \operatorname{epi} g' + \eta\mathbb{S}$ if and only if it satisfies

$$\operatorname{epi} g \cap \big((\rho I\!\!B \cap \operatorname{lev}_\rho g) \times [-\rho, \infty)\big) \subset \operatorname{epi} g' + \eta(I\!\!B \times [-1, \infty)),$$

where $I\!\!B$ is the unit ball at the origin of $I\!\!R^n$. This is also equivalent to

$$\operatorname{epi}\big(\max\{g, -\rho\} + \delta_{\rho \boldsymbol{B} \cap \operatorname{lev}_\rho g}\big) \subset \operatorname{epi} g' + \operatorname{epi}(\delta_{\eta \boldsymbol{B}} - \eta),$$

where δ_S is the function on $I\!\!R^n$ that equals 0 on $S \subset I\!\!R^n$ and ∞ otherwise. In view of [29, Exercise 1.28], we observe that

$$\operatorname{epi} g' + \operatorname{epi}(\delta_{\eta\boldsymbol{B}} - \eta) = \operatorname{epi} g'_\eta, \text{ with } g'_\eta(x) = \min_{y \in \boldsymbol{B}(x,\eta)} g'(y) - \eta.$$

Thus, $\operatorname{epi} g \cap \rho\mathbb{S} \subset \operatorname{epi} g' + \eta\mathbb{S}$ holds if and only if $\min_{y\in\boldsymbol{B}(x,\eta)} g'(y) \leq \max\{g(x), -\rho\} + \eta$ for all $x \in \rho I\!\!B \cap \operatorname{lev}_\rho g$. The conclusion follows after an identical argument with the roles of g and g' reversed. □

We next show the implication of these results and definitions for estimates of minsup-points and minsup-values. Although these estimates could have been stated in terms of the lop-distance, simplifications accrue for working directly with the auxiliary quantity $\hat{d\!l}^{\mathrm{e}}_\rho(h, h')$. In view of Theorem 5.2, the difference is anyhow minor. We state the result for general functions and specialize to bifunctions in a corollary. The result resembles [29, Theorem 7.64], but uses different assumptions. Let $I\!\!R_+ = [0, \infty)$.

5.4 Theorem (estimates of optimal values and solutions) *Suppose that $g, g' \in$ lsc-fcns$(I\!\!R^n)$ and $\rho \geq 0$ is such that*

$$\rho \geq \inf g \geq -\rho \;\; \text{and} \;\; \operatorname{argmin} g \cap \rho I\!\!B \neq \emptyset,$$

with an identical condition for g'. Then,

$$|\inf g' - \inf g| \leq \hat{d\!l}^{\mathrm{e}}_\rho(g, g').$$

If in addition there exists an increasing and continuous function $\psi_g : I\!\!R_+ \to I\!\!R_+$, with $\psi_g(0) = 0$, such that[11]

$$g(x) - \inf g \geq \psi_g\big(\operatorname{dist}(x, \operatorname{argmin} g)\big) \text{ for all } x \in I\!\!R^n,$$

then

$$\Big(\operatorname{argmin} g' \cap \rho I\!\!B\Big) \subset \operatorname{argmin} g + \Big(\hat{d\!l}^{\mathrm{e}}_\rho(g, g') + \psi_g^{-1}\big(2\hat{d\!l}^{\mathrm{e}}_\rho(g, g')\big)\Big) I\!\!B.$$

Proof. Let $\eta \in (\hat{d\!l}^{\mathrm{e}}_\rho(g, g'), \infty)$ be arbitrary. Then, for all $x \in \rho I\!\!B \cap \operatorname{lev}_\rho g$,

$$\inf g' \leq \min_{y \in \boldsymbol{B}(x,\eta)} g'(y) \leq \max\{g(x), -\rho\} + \eta. \tag{6}$$

Let $x \in \operatorname{argmin} g \cap \rho I\!\!B$. Since $\inf g \leq \rho$, $x \in \operatorname{lev}_\rho g$. In view of (6), we find that

$$\inf g' \leq \max\{g(x), -\rho\} + \eta = \max\{\inf g, -\rho\} + \eta \leq \inf g + \eta.$$

[11] The distance between a point $x \in I\!\!R^n$ and a set $A \subset I\!\!R^n$, $\operatorname{dist}(x, A)$, is based on the Euclidean distance.

After letting η tend to its lower limit, we obtain that $\inf g' \le \inf g + \hat{d\!l}^{\mathrm{e}}_\rho(g,g')$. The first result follows after a replication of these arguments with the roles of g and g' reversed.

Next, we establish the solution estimates. Again, let $\eta \in (\hat{d\!l}^{\mathrm{e}}_\rho(g,g'), \infty)$ be arbitrary. For all $x \in \rho I\!\!B \cap \mathrm{lev}_\rho\, g'$,

$$\min_{y\in \boldsymbol{B}(x,\eta)} g(y) \le \max\{g'(x), -\rho\} + \eta. \tag{7}$$

In view of the property of ψ_g and the fact that $\inf g' \ge -\rho$, we find that for $x \in \rho I\!\!B \cap \mathrm{lev}_\rho\, g'$,

$$\begin{aligned} g'(x) - \inf g + \eta = \max\{g'(x), -\rho\} + \eta - \inf g &\ge \min_{y\in \boldsymbol{B}(x,\eta)} g(y) - \inf g \\ &\ge \min_{y\in \boldsymbol{B}(x,\eta)} \psi_g\big(\,\mathrm{dist}\,\big(y, \mathrm{argmin}\, g\big)\big). \end{aligned}$$

From above, $\inf g' - \inf g \le \eta$. Thus, for $x \in \mathrm{argmin}\, g' \cap \rho I\!\!B$, which of course implies that $x \in \mathrm{lev}_\rho\, g'$, we have that

$$\begin{aligned} 2\eta &\ge \inf g' - \inf g + \eta = g'(x) - \inf g + \eta \\ &\ge \min_{y\in \boldsymbol{B}(x,\eta)} \psi_g\big(\,\mathrm{dist}\,\big(y, \mathrm{argmin}\, g\big)\big) \\ &\ge \psi_g\big(\,\min\nolimits_{y\in \boldsymbol{B}(x,\eta)} \mathrm{dist}\,\big(y, \mathrm{argmin}\, g\big)\big), \end{aligned}$$

where the last inequality follows from the increasing property of ψ_g. Therefore, we have that

$$\min_{y\in \boldsymbol{B}(x,\eta)} \mathrm{dist}\,\big(y, \mathrm{argmin}\, g\big) \le \psi_g^{-1}(2\eta).$$

There exists an $\bar{x} \in I\!\!B(x,\eta)$ such that $\mathrm{dist}(\bar{x}, \mathrm{argmin}\, g) = \min_{y\in \boldsymbol{B}(x,\eta)} \mathrm{dist}(y, \mathrm{argmin}\, g)$. These facts then imply that

$$\mathrm{dist}(x, \mathrm{argmin}\, g) \le \mathrm{dist}(\bar{x}, \mathrm{argmin}\, g) + \|\bar{x} - x\| \le \min_{y\in \boldsymbol{B}(x,\eta)} \mathrm{dist}\,\big(y, \mathrm{argmin}\, g\big) + \eta \le \psi_f^{-1}(2\eta) + \eta.$$

Since ψ_g^{-1} is continuous, the conclusion follows by letting η tend to $\hat{d\!l}^{\mathrm{e}}_\rho(g,g')$. □

An intuitive understanding of the conditioning function ψ_g might stem from the simple observation that if $g(x) = \|x\|^2$, then we can select $\psi_g(\tau) = \tau^2$ and $\psi_g^{-1}(\eta) = \sqrt{\eta}$. Direct application of the previous theorem to sup-projections of bifunctions yields the following corollary. The strictness of the above bound is established by considering the instance $g(x) = 0$ for $x = 0$ and otherwise $g(x) = \infty$, and $g(x) = \eta$ for $x = \eta$ and otherwise $g'(x) = \infty$, where $\eta > 0$. Then, $\hat{d\!l}^{\mathrm{e}}_\rho(g,g') = \eta$ for large enough ρ. In this case, the conditioning function ψ_g can be arbitrarily steep. Clearly, the difference between both optimal values and optimal solutions is η.

5.5 Corollary (estimates of minsup-points and values) *Let $f_{XY}, f'_{X'Y'}$ be bifunctions, with sup-projections $h, h' \in$ lsc-fcns$(I\!\!R^n)$, and let $\rho \ge 0$ be such that*

$$\rho \ge \inf\nolimits_X \sup\nolimits_Y f \ge -\rho \ \text{ and } \ \mathrm{argmin}_X \sup\nolimits_Y f \cap \rho I\!\!B \ne \emptyset,$$

with an identical requirement on $f'_{X'Y'}$. Then,

$$|\inf\nolimits_{X'} \sup\nolimits_{Y'} f' - \inf\nolimits_X \sup\nolimits_Y f| \le \hat{d\!l}^{\mathrm{e}}_\rho(h,h').$$

If in addition there exists an increasing and continuous function $\psi_h : I\!R_+ \to I\!R_+$, with $\psi_h(0) = 0$, such that

$$h(x) - \inf_X \sup_Y f \geq \psi_h\big(\operatorname{dist}(x, \operatorname{argmin}_X \sup_Y f)\big) \text{ for all } x \in I\!R^n,$$

then

$$\Big(\operatorname{argmin}_{X'} \sup_{Y'} f' \cap \rho I\!B\Big) \subset \operatorname{argmin}_X \sup_Y f + \Big(\hat{d\!l}^{\,\mathrm{e}}_\rho(h, h') + \psi_h^{-1}\big(2\hat{d\!l}^{\,\mathrm{e}}_\rho(h, h')\big)\Big) I\!B.$$

Corollary 5.5 is significant as it ties directly the distances between minsup-points and minsup-values of two bifunctions to the corresponding auxiliary quantity $\hat{d\!l}^{\,\mathrm{e}}_\rho$ and, through Theorem 5.2, the lop-distance. In fact, the lop-distance was constructed with this goal in mind.

A useful estimate of $\hat{d\!l}^{\,\mathrm{e}}_\rho$ is given next. We say that a function $g : I\!R^n \to I\!R$ is Lipschitz continuous with modulus $\kappa : I\!R_+ \to I\!R_+$ if

$$|g(x) - g(x')| \leq \kappa(\rho)\|x - x'\| \text{ for all } x, x' \in \rho I\!B.$$

A component towards an estimate is a distance between two nonempty closed subsets C, D of $I\!R^k$ given by

$$\hat{d\!l}_\rho(C, D) := \{\inf \eta \geq 0 \;:\; C \cap \rho I\!B \subset D + \eta I\!B, D \cap \rho I\!B \subset C + \eta I\!B\}, \text{ for } \rho \geq 0,$$

where $I\!B$ is the Euclidean ball in $I\!R^k$. This quantity is closely related to $\hat{d\!l}^{\,\mathrm{e}}_\rho$, but deals with arbitrary nonempty closed set, not only epi-graphs, and also rely on the usual Euclidean norm and not $\|\cdot\|_{\mathbb{S}}$. The next proposition improves on [29, Exercise 7.62].

5.6 Proposition *Suppose that $C, C' \subset I\!R^n$ are nonempty closed sets and $g_0, g_0' : I\!R^n \to I\!R$ are Lipschitz continuous with common modulus $\kappa : I\!R_+ \to I\!R_+$. Then, for $\rho \in [0, \infty)$ and $\rho' \geq \rho + \hat{d\!l}_\rho(C, C')$, the functions*[12] *$g = g_0 + \delta_C$ and $g' = g_0' + \delta_{C'}$ has*

$$\hat{d\!l}^{\,\mathrm{e}}_\rho(g, g') \leq \max_{x \in \rho \boldsymbol{B}} |g_0(x) - g_0'(x)| + \max\{1, \kappa(\rho')\}\hat{d\!l}_\rho(C, C').$$

Proof. Set $\eta = \max_{x \in \rho \boldsymbol{B}} |g_0(x) - g_0'(x)| + \max\{1, \kappa(\rho')\}\hat{d\!l}_\rho(C, C')$. First, we establish that

$$\min_{y \in \boldsymbol{B}(x,\eta)} g'(y) \leq \max\{g(x), -\rho\} + \eta \text{ for } x \in \rho I\!B \cap \operatorname{lev}_\rho g.$$

Suppose that $x \in \rho I\!B \cap \operatorname{lev}_\rho g$, which of course implies that $x \in C$. There exists a $y \in C'$ such that $\|x - y\| = \operatorname{dist}(x, C')$. Thus,

$$\hat{d\!l}_\rho(C, C') \geq \inf\{\eta' \geq 0 \;:\; C \cap \rho I\!B \subset C' + \eta' I\!B\} \geq \operatorname{dist}(x, C') = \|x - y\|$$

and therefore $\|x - y\| \leq \eta$. Moreover, $\|y\| \leq \|x\| + \|x - y\| \leq \rho + \hat{d\!l}_\rho(C, C') \leq \rho'$. These facts and the Lipschitz continuity of g_0' on $\rho' I\!B$ imply that

$$\begin{aligned}
\min_{y' \in \boldsymbol{B}(x,\eta)} g'(y') &\leq g'(y) = g_0'(y) = g_0'(y) - g_0'(x) + g_0'(x) - g_0(x) + g_0(x) \\
&\leq \kappa(\rho')\|x - y\| + \sup_{\rho \boldsymbol{B}} |g_0 - g_0'| + \max\{g(x), -\rho\} \\
&\leq \max\{g(x), -\rho\} + \eta,
\end{aligned}$$

[12]For a set $C \subset I\!R^n$, the function δ_C is zero on C and ∞ elsewhere.

which establishes the claim. Second, a parallel argument with the roles of g and g' reversed, leads to the conclusion. $\square$

Although one can apply the previous result directly for sup-projections in place of the functions g and g', we develop a more detailed result under additional assumptions. We limit the scope to $\mathcal{Y} = \mathbb{R}^m$; see [3] for related estimates on normed linear spaces. Up to now, we have only needed that bifunctions are defined and finite-valued for every $x \in X$ and $y \in Y(x)$. However, to get relatively simple expressions, we now assume that bifunctions are defined and finite-valued on $\mathbb{R}^n \times \mathbb{R}^m$. For a function $g : \mathbb{R}^k \to \mathbb{R}$ and nonempty $Z \subset \mathbb{R}^k$, with a slight abuse of notation, we denote by $g : Z \to \mathbb{R}$ the restriction of g to Z, with the usual interpretation that $g : Z \to \mathbb{R}$ is then assigned the value ∞ outside Z.

We say that $f : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ is *marginally Lipschitz continuous with moduli* $\kappa_x : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ and $\kappa_y : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ if for every $\rho, \lambda \geq 0$,

$$|f(x,y) - f(x,y')| \leq \kappa_y(\rho,\lambda)\|y - y'\| \text{ for all } x \in \rho\mathbb{B},\ y, y' \in \lambda\mathbb{B}$$
$$|f(x,y) - f(x',y)| \leq \kappa_x(\rho,\lambda)\|x - x'\| \text{ for all } x, x' \in \rho\mathbb{B},\ y \in \lambda\mathbb{B}.$$

We say that a set-valued mapping $Y : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *uniformly* $\hat{d\kern-0.3em l}_\lambda$*-Lipschitz continuous with modulus* $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$ if for all $\rho \geq 0$ and $\lambda \geq 0$,

$$\hat{d\kern-0.3em l}_\lambda\big(Y(x), Y(x')\big) \leq \gamma(\rho)\|x - x'\| \text{ for all } x, x' \in \rho\mathbb{B}.$$

This is a strong property, but it holds with $\gamma(\rho) = 0$ if Y is independent of x.

5.7 Theorem (bound under Lipschitz continuity) *Suppose that $f, f' : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ are marginally Lipschitz continuous with common moduli $\kappa_x, \kappa_y : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$, the set-valued mappings $Y, Y' : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ are closed- and nonempty-valued and uniformly $\hat{d\kern-0.3em l}_\lambda$-Lipschitz continuous with common modulus $\gamma : \mathbb{R}_+ \to \mathbb{R}_+$, and $X, X' \subset \mathbb{R}^n$ are nonempty closed sets. Moreover, suppose that $\{y \in Y(x) : f(x,y) \geq \alpha\}$ and $\{y \in Y'(x) : f'(x,y) \geq \alpha\}$ are bounded for all $\alpha \in \mathbb{R}$ and $x \in \mathbb{R}^n$.*

For a fixed $\rho \geq 0$, suppose that for some $\rho^* \geq \rho + \hat{d\kern-0.3em l}_\rho(X, X')$, $\lambda^* \geq 0$, and $\sigma < \infty$, we have that

(i) $\hat{d\kern-0.3em l}_{\lambda^*}(Y(x), Y'(x)) \leq \sigma$ *for all* $x \in \rho\mathbb{B}$

(ii) $\operatorname{argmax}_{y \in Y(x)} f(x,y) \cap \lambda^*\mathbb{B} \neq \emptyset$ *and* $\operatorname{argmax}_{y \in Y'(x)} f'(x,y) \cap \lambda^*\mathbb{B} \neq \emptyset$ *for all* $x \in \rho^*\mathbb{B}$

(iii) $\sup_{y \in Y(x)} f(x,y), \sup_{y \in Y'(x)} f'(x,y) \in [-\lambda^*, \lambda^*]$ *for all* $x \in \rho^*\mathbb{B}$.

Then, the sup-projections h and h' corresponding to f_{XY} and $f'_{X'Y'}$, respectively, satisfy

$$\hat{d\kern-0.3em l}^{\,\mathrm{e}}_\rho\big(h : X \to \mathbb{R}, h' : X' \to \mathbb{R}\big) \leq \max_{x \in \rho\mathbb{B},\ y \in \lambda^*\mathbb{B}} \big|f(x,y) - f'(x,y)\big|$$
$$+ \bar{\kappa}_x \hat{d\kern-0.3em l}_\rho(X, X') + \bar{\kappa}_y \sup_{x \in \rho\mathbb{B}} \hat{d\kern-0.3em l}_{\lambda^*}\big(Y(x), Y'(x)\big),$$

where $\bar{\kappa}_x = \max\{1, \kappa_x(\rho^*, \lambda^*) + \max\{1, \kappa_y(\rho^*, \lambda^* + \gamma(\rho^*)\rho^*)\}\gamma(\rho^*)\}$ *and* $\bar{\kappa}_y = \max\{1, \kappa_y(\rho, \lambda^* + \sigma)\}$.

Proof. For any ρ and λ, let $\kappa'_y(\rho, \lambda) = \max\{1, \kappa_y(\rho, \lambda)\}$. The result is a consequence of repeated applications of Proposition 5.6 and Theorem 5.4. Fix $\rho \geq 0$ and let $\rho^* \geq \rho + \hat{d\kern-0.3em l}_\rho(X, X')$ be such that assumptions hold. We start with an application of Proposition 5.6. For any $x \in \rho^* I\!B$, $Y(x)$ is closed and nonempty, and $f(x,\cdot) : I\!R^m \to I\!R$ satisfies

$$|f(x,y) - f(x,y')| \le \kappa_y(\rho^*,\lambda)\|y - y'\| \text{ for any } \lambda \ge 0 \text{ and } y,y' \in \lambda I\!B.$$

Thus, by Proposition 5.6, for $x,x' \in \rho^* I\!B$ and $\lambda \ge 0$,

$$\begin{aligned} \hat{d\!l}^{\mathrm{e}}_\lambda\Big(-f(x,\cdot) : Y(x) \to I\!R,\ -f(x',\cdot) : Y(x') \to I\!R\Big) \le \max\nolimits_{y\in\lambda \boldsymbol{B}} \big|f(x,y) - f(x',y)\big| \\ + \kappa_y'(\rho^*,\lambda')\hat{d\!l}_\lambda\big(Y(x), Y(x')\big), \end{aligned}$$

when $\lambda' \ge \lambda + \hat{d\!l}_\lambda(Y(x), Y(x'))$. For $x,x' \in \rho^* I\!B$ and $\lambda \ge 0$, the Lipschitz continuity of Y implies that $\hat{d\!l}_\lambda(Y(x), Y(x')) \le \gamma(\rho^*)\|x - x'\| \le \gamma(\rho^*)\rho^*$. These facts and the Lipschitz continuity of $f(\cdot,y)$ establish that for $x,x' \in \rho^* I\!B$ and $\lambda \ge 0$,

$$\hat{d\!l}^{\mathrm{e}}_\lambda\Big(-f(x,\cdot) : Y(x) \to I\!R,\ -f(x',\cdot) : Y(x') \to I\!R\Big) \le \Big[\kappa_x(\rho^*,\lambda)+\kappa_y'\big(\rho^*,\lambda+\gamma(\rho^*)\rho^*\big)\gamma(\rho^*)\Big]\|x-x'\|. \quad (8)$$

An identical argument gives the same result for f' and Y'.

We next apply Theorem 5.4, reoriented towards maximization instead of minimization, and find that for $x,x' \in \rho^* I\!B$,

$$\big|\sup\nolimits_{y\in Y(x)} f(x,y) - \sup\nolimits_{y\in Y(x')} f(x',y)\big| \le \hat{d\!l}^{\mathrm{e}}_{\lambda^*}\big(-f(x,\cdot) : Y(x) \to I\!R, -f(x',\cdot) : Y(x') \to I\!R\big).$$

Utilizing the bound (8) on the right-hand side gives that

$$\big|h(x) - h(x')\big| \le \Big[\kappa_x(\rho^*,\lambda^*) + \kappa_y'\big(\rho^*,\lambda^* + \gamma(\rho^*)\rho^*\big)\gamma(\rho^*)\Big]\|x - x'\| \text{ for all } x,x' \in \rho^* I\!B.$$

An identical argument establishes the same Lipschitz property for h' on $\rho^* I\!B$.

We next apply Proposition 5.6 for a second time. Under the stated assumptions h,h' are finite valued on $I\!R^n$ because, for all $x \in I\!R^n$, $Y(x)$ and $Y'(x)$ are nonempty and closed, and $f(x,\cdot) : Y(x) \to I\!R$ and $f'(x,\cdot) : Y'(x) \to I\!R$ are superlevel-bounded. Moreover, h,h' satisfy the previously established Lipschitz property on $\rho^* I\!B$. Thus, by Proposition 5.6,

$$\begin{aligned} \hat{d\!l}^{\mathrm{e}}_\rho\big(h : X \to I\!R, h' : X' \to I\!R\big) \le \max\nolimits_{x\in\rho \boldsymbol{B}} |h(x) - h'(x)| \\ + \max\Big\{1, \kappa_x(\rho^*,\lambda^*) + \kappa_y'\big(\rho^*,\lambda^* + \gamma(\rho^*)\rho^*\big)\gamma(\rho^*)\Big\}\hat{d\!l}_\rho(X,X'). \end{aligned} \quad (9)$$

It remains to bound $|h(x)-h'(x)|$ for $x \in \rho I\!B$. An intermediate step is a third application of Proposition 5.6. For $x \in \rho I\!B$, by Proposition 5.6,

$$\begin{aligned} \hat{d\!l}^{\mathrm{e}}_\lambda\big(-f(x,\cdot) : Y(x) \to I\!R, f'(x,\cdot) : Y'(x) \to I\!R\big) \le \max\nolimits_{y\in\lambda^* \boldsymbol{B}} \big|f(x,y) - f'(x,y)\big| \\ + \kappa_y'(\rho,\lambda^* + \sigma)\hat{d\!l}_{\lambda^*}\big(Y(x), Y'(x)\big). \end{aligned} \quad (10)$$

Another application of Theorem 5.4 establishes the bound

$$\left|\sup_{y\in Y(x)} f(x,y) - \sup_{y\in Y'(x)} f'(x,y)\right| \le \hat{d\!l}^{\mathrm{e}}_{\lambda^*}\big(-f(x,\cdot) : Y(x) \to I\!R, -f'(x,\cdot) : Y'(x) \to I\!R\big) \text{ for } x \in \rho I\!B.$$

Combining this result with (10), we obtain that

$$\max_{x\in\rho\boldsymbol{B}} |h(x) - h'(x)| \leq \max_{x\in\rho\boldsymbol{B},\ y\in\lambda^*\boldsymbol{B}} \big|f(x,y) - f'(x,y)\big| + \kappa'_y(\rho, \lambda^* + \sigma) \sup_{x\in\rho\boldsymbol{B}} \hat{d\!l}_{\lambda^*}\big(Y(x), Y'(x)\big).$$

In view of this bound and (9), the conclusion follows. □

When Y, Y' are independent of x, the Lipschitz constants are uniform, and other assumptions, the result simplifies considerably.

5.8 Corollary *Suppose that $f, f' : I\!R^n \times I\!R^m \to I\!R$ are marginally Lipschitz continuous, with common moduli κ_x, κ_y independent of ρ, λ; $X, X' \subset I\!R^n$ are nonempty closed sets; and $Y, Y' \subset I\!R^m$ are nonempty compact sets.*

For $\rho \geq 0$, suppose that for some $\rho^* \geq \rho + \hat{d\!l}_\rho(X, X')$ and $\lambda^* \geq 0$, we have that $\operatorname{argmax}_{y\in Y} f(x,y) \cap \lambda^* I\!B \neq \emptyset$, $\operatorname{argmax}_{y\in Y'} f'(x,y) \cap \lambda^* I\!B \neq \emptyset$, and $\max_{y\in Y} f(x,y), \max_{y\in Y'} f'(x,y) \in [-\lambda^*, \lambda^*]$ for all $x \in \rho^* I\!B$. Then,

$$\begin{aligned}\hat{d\!l}{}^{\mathrm{e}}_\rho(h : X \to I\!R, h' : X' \to I\!R) \leq\ & \max_{x\in\rho\boldsymbol{B},\ y\in\lambda^*\boldsymbol{B}} |f(x,y) - f'(x,y)| \\ & + \max\{1, \kappa_x\}\hat{d\!l}_\rho(X, X') + \max\{1, \kappa_y\}\hat{d\!l}_{\lambda^*}(Y, Y').\end{aligned}$$

5.3 Estimates in Robust and Risk-Averse Optimization

We end the paper with two simple illustrations of estimates of the distance between minsup-values. The examples do not represent the limit of the technology developed here. In fact, the examples are deliberately selected to be simple. Certainly, in these cases, similar results can be obtained using other means.

Example 9 (cont.): For $\lambda > 0$ sufficiently large and $\alpha, \beta \in [0, 1)$, it is easy to show that

$$\hat{d\!l}_\lambda(Y_\alpha, Y_\beta) \leq \frac{\sqrt{J}|\alpha - \beta|}{(1-\alpha)(1-\beta)}.$$

In view of Corollaries 5.5 and 5.8, if X is compact, then for $0 \leq \alpha \leq \beta < 1$

$$\inf_X \sup_{Y_\beta} f \geq \inf_X \sup_{Y_\alpha} f \geq \inf_X \sup_{Y_\beta} f - \max\left\{1, \sup_{x\in X}\sqrt{\sum_{j=1}^J \psi(x, \xi^j)^2}\right\} \frac{\sqrt{J}(\beta - \alpha)}{(1-\beta)^2}.$$

This expression provides a bound on the price of robustness. For example, if β is fixed at a value near 1, implying a highly risk-averse perspective, and α approaches β from below, then the lower bound on the minsup-value for f_{XY_α} increases at a linear rate towards the "high" minsup-value of f_{XY_β}. This result can then be used to assess the price associated increasing robustness from a low α to a higher α.

Example 10 (cont.): We return to the robust investment problem, but simplify the uncertainty set by letting

$$\Xi_u = \left\{\xi \in I\!R^n \ :\ \sum_{i=1}^n \left(\frac{\xi_i - \bar{\xi}_i}{\sigma_i}\right)^2 \leq u^2\right\}, \text{ with } u \geq 0,$$

be independent of x. Let $\sigma = \max_{i=1,\dots,n} \sigma_i$. Then, it easy to see that for $u, v \geq 0$ and sufficiently large $\lambda \geq 0$,

$$\hat{d\!l}_\lambda(\Xi_u, \Xi_v) = \sigma\, |u - v|\, .$$

Thus, in view of Corollary 5.5 and the fact that $|\langle \xi, x\rangle - \langle \xi', x\rangle| \leq \|x\| \|\xi - \xi'\|$, we find that the corresponding sup-projections $h_u = \sup_{\xi \in \Xi_u} \langle -\xi, \cdot\rangle$ and $h_v = \sup_{\xi\in\Xi_v} \langle -\xi, \cdot\rangle$ satisfy

$$\hat{d\!l}{}^{\mathrm{e}}_\rho (h_u,\ h_v) \leq \max\{1, \rho\}\sigma\, |u - v|\, .$$

Consequently, in view of Theorem 5.4, $\min_{x\in X} \sup_{\xi\in\Xi_v} \langle -\xi, x\rangle$ is no lower than $\min_{x\in X} \sup_{\xi\in\Xi_u} \langle -\xi, x\rangle - \max\{1, \rho\}\sigma|u - v|$ for $u \geq v$.

Acknowledgement. This work in supported in parts by DARPA under HR0011-14-1-0060 as well as the U. S. Army Research Laboratory and the U. S. Army Research Office under grant numbers 00101-80683, W911NF-10-1-0246, and W911NF-12-1-0273.

Appendix

Proof of Theorem 3.7. Since $d\!l^{\mathrm{h}}(F, G) = d\!l^{\mathrm{e}}(-F, -G)$, we follow an argument that is similar to that of Theorem 5.2. Utilizing the facts that $\mathrm{dist}(0, \mathrm{epi}(-F)) = \mathrm{dist}(0, \mathrm{epi}(-G)) = 0$, and F and G are both bounded between 0 and 1, a line of arguments along those in the proof of Theorem 5.2 leads to

$$e^{-\rho}\hat{d\!l}{}^{\mathrm{e}}_\rho(-F, -G) \leq d\!l^{\mathrm{e}}(-F, -G) \leq e^{-\rho} + (1 - e^{-\rho})\hat{d\!l}{}^{\mathrm{e}}_{2\rho}(-F, -G).$$

Thus, we only need to construct a lower bound on $\hat{d\!l}{}^{\mathrm{e}}_\rho(-F, -G)$ and an upper bound on $\hat{d\!l}{}^{\mathrm{e}}_{2\rho}(-F, -G)$. In both cases, we utilize Proposition 5.3.

First, we consider the lower bound. Since $F, G \leq 1$, a $\rho \geq 1$ simplifies the alternative expression for $\hat{d\!l}{}^{\mathrm{e}}_\rho(-F, -G)$ in Proposition 5.3 to

$$\begin{aligned}\hat{d\!l}{}^{\mathrm{e}}_\rho(-F, -G) = \inf\Big\{\eta \geq 0 \ :\ &\min_{\boldsymbol{B}(\xi,\eta)} -G \leq -F(\xi) + \eta \\ &\min_{\boldsymbol{B}(\xi,\eta)} -F \leq -G(\xi) + \eta,\ \text{for all } \xi \in \rho I\!\!B\Big\}.\end{aligned}$$

Replacing the minimization over a ball by minimization over the smallest hypercube containing the ball, we obtain a relaxation of the infimum problem over η. Due to the monotonicity of F and G, the minimization over the hypercube is attained at a particular vertex. Hence, for $\rho \geq 1$,

$$\begin{aligned}\hat{d\!l}{}^{\mathrm{e}}_\rho(-F, -G) \geq \inf\Big\{\eta \geq 0 \ :\ &G(\xi + \eta\mathbf{1}) + \eta \geq F(\xi) \\ &F(\xi + \eta\mathbf{1}) + \eta \geq G(\xi),\ \text{for all } \xi \in \rho I\!\!B\Big\}.\end{aligned}$$

Second, we consider an upper bound on $\hat{d\!l}{}^{\mathrm{e}}_{2\rho}(-F, -G)$. Similar to the case for the lower bound, because $F, G \leq 1$, a $\rho \geq 1/2$ simplifies the expression for $\hat{d\!l}{}^{\mathrm{e}}_{2\rho}(-F, -G)$ to

$$\begin{aligned}\hat{d\!l}{}^{\mathrm{e}}_{2\rho}(-F, -G) = \inf\Big\{\eta \geq 0 \ :\ &\min_{\boldsymbol{B}(\xi,\eta)} -G \leq -F(\xi) + \eta \\ &\min_{\boldsymbol{B}(\xi,\eta)} -F \leq -G(\xi) + \eta,\ \text{for all } \xi \in 2\rho I\!\!B\Big\}.\end{aligned}$$

Replacing the minimization over a ball by minimization over the largest hypercube contained in the ball, we obtain a restrictions of the infimum problem over η. Due to the monotonicity of F and G, the minimization over the hypercube is attained at a particular vertex. Hence, for $\rho \geq 1/2$,

$$\hat{d\!l}^{\mathrm{e}}_{2\rho}(-F,-G) \leq \inf \Big\{ \eta \geq 0 \;:\; G(\xi + \eta \mathbf{1}/\sqrt{m}) + \eta \geq F(\xi) \\ F(\xi + \eta \mathbf{1}/\sqrt{m}) + \eta \geq G(\xi), \text{ for all } \xi \in 2\rho I\!B \Big\}.$$

Denoting the lower bounding and upper bounding expressions by $\underline{\eta}(\rho)$ and $\overline{\eta}(\rho)$, respectively, yields two first inequalities. Letting $\rho \to \infty$ in the upper bound we obtain the last inequality. $\square$

References

[1] P. Artzner, F. Delbaen, J.-M. Eber, and D. Heath. Coherent measures of risk. *Mathematical Finance*, 9:203–227, 1999.

[2] H. Attouch and R. J-B Wets. Convergence des points min/sup et de points fixes. *Comptes Rendus de l'Académie des Sciences de Paris*, 296:657–660, 1983.

[3] H. Attouch and R. J-B Wets. Quantitative stability of variational systems: I. the epigraphical distance. *Transactions of the American Mathematical Society*, 328(2):695–729, 1991.

[4] G. Beer. *Topologies on Closed and Closed Convex Sets*, volume 268 of *Mathematics and its Applications*. Kluwer, 1992.

[5] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski. *Robust Optimization*. Princeton University Press, 2009.

[6] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Operations Research Letters*, 25:1–13, 1999.

[7] D. Bertsimas, D. B. Brown, and C. Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[8] P. Billingsley. *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. Wiley, 1995.

[9] J. R. Birge and F. Louveaux. *Introduction to Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer, 2. edition, 2011.

[10] G. C. Calafiore. Ambiguous risk measures and optimal robust portfolios. *SIAM Journal on Optimization*, 18:853–877, 2007.

[11] R. M. Dudley. *Real Analysis and Probability*. Cambridge University Press, 2004.

[12] L. El Ghaoui and G. C. Calafiore. *Optimization Models*. Cambridge University Press, 2014.

[13] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. de Gruyter, 3. edition, 2011.

[14] M. A. Goberna and M. A. Lopez. *Semi-Infinite Programming: Recent Advances*. Nonconvex optimization and applications. Springer, 2001.

[15] A. Jofre and R. J-B Wets. Variational convergence of bivariate functions: Lopsided convergence. *Mathematical Programming B*, 116:275–295, 2009.

[16] A. Jofre and R. J-B Wets. Variational convergence of bifunctions: Motivating applications. *SIAM J. Optimization*, 24(4):1952–1979, 2014.

[17] T. W. Jonsbråten, R. J-B Wets, and D. L. Woodruff. A class of stochastic programs with decision dependent random elements. *Annals of Operations Research*, 82:83–106, 1998.

[18] A. J. King and S. W. Wallace. *Modeling with Stochastic Programming*. Springer Series in Operations Research and Financial Engineering. Springer, 2013.

[19] S. Mehrotra and D. Papp. A cutting surface algorithm for semi-infinite convex programming, with an application to distributionally robust optimization. *SIAM J. Optimization*, 24(4):1670–1697, 2014.

[20] G. Pflug, A. Pichler, and D. Wozabal. The $1/n$ investment strategy is optimal under high model ambiguity. *Journal of Banking & Finance*, 36(2):410–417, 2012.

[21] G. Ch. Pflug and W. Römish. *Modeling, Measuring and Managing Risk*. World Scientific, 2007.

[22] E. Polak. *Optimization. Algorithms and Consistent Approximations*, volume 124 of *Applied Mathematical Sciences*. Springer, 1997.

[23] R. T. Rockafellar and J. O. Royset. On buffered failure probability in design and optimization of structures. *Reliability Engineering & System Safety*, 95:499–510, 2010.

[24] R. T. Rockafellar and J. O. Royset. Random variables, monotone relations, and convex analysis. *Mathematical Programming B*, 148(1):297–331, 2014.

[25] R. T. Rockafellar and J. O. Royset. Measures of residual risk with connections to regression, risk tracking, surrogate models, and ambiguity. *SIAM J. Optimization*, 25(2):1179–1208, 2015.

[26] R. T. Rockafellar and J. O. Royset. Superquantile/CVaR risk measures: Second-order theory. *Annals of Operations Research*, OnlineFirst, 2016.

[27] R. T. Rockafellar and S. Uryasev. Optimization of conditional Value-at-Risk. *Journal of Risk*, 2:493–517, 2000.

[28] R. T. Rockafellar and S. Uryasev. The fundamental risk quadrangle in risk management, optimization and statistical estimation. *Surveys in Operations Research and Management Science*, 18:33–53, 2013.

[29] R. T. Rockafellar and R. J-B Wets. *Variational Analysis*, volume 317 of *Grundlehren der Mathematischen Wissenschaft*. Springer, 3rd printing-2009 edition, 1998.

[30] J. O. Royset and E. Polak. On the use of Augmented Lagrangians in the solution of generalized semi-infinite min-max problems. *Computational Optimization and Applications*, 31(2):173–192, 2005.

[31] J. O. Royset, E. Polak, and A. Der Kiureghian. Adaptive approximations and exact penalization for the solution of generalized semi-infinite min-max problems. *SIAM J. Optimization*, 14(1):1–34, 2003.

[32] J. O. Royset and R. J-B Wets. Multivariate epi-splines and evolving function identification problems. *Set-Valued and Variational Analysis*, OnlineFirst, 2016.

[33] J. O. Royset and R. J-B Wets. On lopsided convergence and its quantifications. In review, 2016.

[34] J. O. Royset and R. J-B Wets. Optimality functions and lopsided convergence. *J. Optimization Theory and Applications*, OnlineFirst, 2016.

[35] A. Ruszczynski and A. Shapiro. Optimization of convex risk functions. *Mathematics of Operations Research*, 31:433–452, 2006.

[36] G. Salinetti and R. J-B Wets. On the convergence in distribution of measurable multifunctions (random sets), normal integrands, stochastic processes and stochastic infima. *Mathematics of Operations Research*, 11(3):385–419, 1986.

[37] G. Salinetti and R. J-B Wets. On the hypo-convergence of probability measures. In *Optimization and Related Fields, Proc., Erice 1984, Lecture Notes in Mathematics 1190*, pages 371–395. Springer, 1986.

[38] S. Sarykalin, G. Serraino, and S. Uryasev. Value-at-Risk vs. Conditional Value-at-Risk in risk management and optimization. In Z.-L. Chen and S. Raghavan, editors, *Tutorials in Operations Research: State-of-the-Art Decision-Making Tools in the Information-Intensive Age*, Tutorials in Operations Research, pages 270–294. INFORMS, Cantonsville, MD, 2008.

[39] A. Shapiro. Distributional robust stochastic programming. *Optimization Online*, 2016.

[40] A. Shapiro, D. Dentcheva, and A. Ruszczynski. *Lectures on Stochastic Programming: Modeling and Theory*. SIAM, 2. edition, 2014.

[41] O. Stein. *Bi-level Strategies in Semi-Infinite Programming*. Kluwer Academic, 2003.